



Forschungszentrum für Umweltpolitik Environmental Policy Research Centre

Green Nudging

A discussion and preliminary evaluation of nudging
as an environmental policy instrument

Nicholas Evans, Stephanie Eickers, Leonie Geene, Marijana
Todorovic, Annika Villmow

FFU-Report 01-2017



Forschungszentrum für Umweltpolitik
Freie Universität Berlin
Fachbereich Politik- und Sozialwissenschaften
Otto-Suhr-Institut für Politikwissenschaft

Publisher

Freie Universität Berlin
Forschungszentrum für Umweltpolitik
Innstraße 22
14195 Berlin
Web: www.fu-berlin.de/ffu

ISSN 1612-3026
FFU-Rep 01-2017

Green Nudging: A discussion and preliminary evaluation of nudging as an environmental policy instrument

This report is the outcome of a research project completed for the Environmental Policy and Planning MA programme at the Freie Universität Berlin. The authors would like to thank Dr. Kirsten Jörgensen for her guidance, encouragement and counsel throughout the research process and Dr. Robert Lepenies for his invaluable insights on the topic and meticulous review. All five authors contributed equally.

Authors

Nicholas Evans

[nickwaevans\[a\]gmail.com](mailto:nickwaevans@gmail.com)

Stephanie Eickers

[stephanie.eickers\[a\]gmail.com](mailto:stephanie.eickers@gmail.com)

Leonie Geene

[leonie.geene\[a\]gmail.com](mailto:leonie.geene@gmail.com)

Marijana Todorovic

[marijana-todorovic\[a\]web.de](mailto:marijana-todorovic@web.de)

Annika Villmow

[annikavillmow\[a\]gmail.com](mailto:annikavillmow@gmail.com)

The Environmental Policy Research Centre (FFU) has been publishing its *FFU reports* since 1993. *FFU reports* encompass both academic papers and policy-oriented studies. All *FFU reports* are either peer-reviewed internally or have been reviewed by other project partners and/or the commissioning client. The views expressed are solely the authors' and do not necessarily reflect those of the FFU.

Abstract

Traditional environmental policy instruments have not always proven successful in fostering environmentally friendly behaviour. The question remains: how can policymakers tackle the attitude-behaviour gap when it comes to pro-environmental choices and sustainable lifestyles? One solution that has emerged is *green nudging*, a new and potentially promising policy tool born of behavioural economics and experimental psychology. This paper contributes to the current discussion surrounding green nudging with an extensive overview of the subject and the establishment of a policy evaluative framework, which, in addition to incorporating the criteria, efficacy and effectiveness, focuses on the commonly neglected dimension of ethics. A preliminary policy evaluation of two types of nudging – defaults and social norms – suggests that while green nudging has proven to be both efficacious in laboratory settings and effective in real life, there remain many ethical concerns that must be accounted for in the implementation of this new policy instrument. Policymakers may be able to use these insights to further develop green nudging as a means to promote pro-environmental behaviour and come a step closer to effective and ethical environmental policymaking.

Contents

1	Introduction	1
2	Background: A theory of nudging	4
2.1	Nudging and its foundations in the behavioural sciences	5
2.1.1	Econs and Humans	5
2.1.2	Dual process theory	8
2.1.3	Heuristics and biases	9
2.1.4	Designing a nudge: Choice architecture	12
2.2	Nudging and libertarian paternalism	14
2.3	Criticism	16
3	Nudging for the environment	17
3.1	The nature of environmental problems	18
3.1.1	The tragedy of the commons	18
3.1.2	Barriers to ecological behaviour	19
3.2	Nudging as an addition to the traditional policy toolkit	20
3.2.1	Traditional environmental policy instruments	21
3.2.2	Nudging in an environmental context	23
3.3	Towards a typology of green nudges	25
3.3.1	Categorising green nudges	25
3.3.2	Proposed typology	26
3.4	Should we nudge for the environment?	30
4	Research design	31
4.1	Conceptualisation and methodological approach	31
4.2	Evaluative framework	32
4.2.1	Efficacy	32
4.2.2	Effectiveness	36
4.2.3	Ethics	39
4.3	Methodological limitations and delimitations	44
4.3.1	Limitations	45
4.3.2	Delimitations	45

5	A policy evaluation framework for nudging	46
5.1	Green Defaults.....	46
5.1.1	Efficacy	47
5.1.2	Effectiveness.....	49
5.1.3	Ethics.....	53
5.2	Green Social Norms	56
5.2.1	Efficacy	56
5.2.2	Effectiveness.....	58
5.2.3	Ethics.....	62
6	Discussion	65
7	Conclusion	68
8	Annex: Overview of green nudges	71
9	References	76

List of figures

Figure 1: A technical definition of nudge	13
Figure 2: Typology of green nudges.....	27

List of tables

Table 1: Traditional environmental policy toolkit.....	21
Table 2: Evaluative framework for green nudging.....	33
Table 3: Three types of validity for the efficacy of policy interventions	34
Table 4: Criteria for an ethical evaluation of green nudges	39
Table 5: Empirical examples of green default nudges	71
Table 6: Empirical examples of green social norm nudges	72

1 Introduction

The concept of nudge originates from the seminal and equally popular book by Richard Thaler and Cass Sunstein of the same name. Less than a decade after the book's publication in 2008, nudging has exploded as a contentious topic of discussion in public policy circles, such as health, taxation, consumer protection, energy and the environment. Apart from sparking nearly a decade's worth of lively debate among academics and practitioners alike, nudging throws a screw into traditional policymaking by challenging the longstanding classical economic paradigm that views humans as perfectly rational beings.

Imagine you are a visitor at the beautiful Yosemite National Park in California. A sign greets you at the trailhead imploring you not to give food to the wildlife. It explains further that each year thousands of hikers feed the park's small fauna, which become dependent on hikers' charity. You begin your hike and intend to obey the sign. Yet, less ecologically conscious people may perceive 'thousands of hikers' as an invitation. Clearly everybody is feeding the wildlife; indeed, what is one additional person? Due to the powerful influence of perceived social norms, the sign's unfortunate wording may lead to an *increase* in the undesirable behaviour! A clever park administrator would instead design a sign that frames feeding wildlife as both uncommon and deviant. Thousands of visitors amount to a drop in the bucket for a park that welcomes nearly four *million* eager tourists annually, and no one wants to belong to the 2-5% of people who harm the local animal populations. Framing the way in which information is displayed can go a long way to solving many environmental problems (Cialdini et al., 2006).

Like the sign in the example above, a nudge, as defined by Thaler and Sunstein (2008), is any "aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives" (p. 6). Thaler and Sunstein (2008) advocate for nudging as a 'libertarian' form of paternalism, claiming that it is legitimate to try to influence peoples' behaviour to make their lives longer, healthier and better – as long as they judge themselves to be better off and their freedom of choice is preserved (p. 5). However, nudging as policy is also unique for its foundation, not in law, political science or economics but rather in the behavioural sciences.

In the past, studies on human behaviour were relegated to disciplines such as sociology or psychology. However, over the past two decades, behavioural science has crossed boundaries and increasingly found entrance into other fields, such as economics and policymaking. In the course of this development, limitations on human rationality have been integrated into economic models and new research areas have mushroomed, such as behavioural economics, behavioural finance and behavioural game theory (Croson & Treich, 2014, p. 335). Through the implementation of nudging in policymaking, citizens can be steered in a desirable direction without having to enforce unpopular rules or regulations. State governments may stand to benefit from the findings of behavioural research by avoiding contentious regulations, bans or taxes and instead gently nudging individuals to save energy, drive more slowly, eat healthier or – in short – to become better citizens.

In 2010, the United Kingdom (UK) was the first to establish a Behavioural Insights Team (BIT) – also called the ‘nudge unit’ – within its Cabinet Office.¹ The purpose of the BIT was to examine how behavioural science findings can tailor public policy to stimulate citizens towards pro-social behaviour, e.g., saving energy or paying taxes. Three years later, following the British example, President Barack Obama signed an executive order establishing the White House Social and Behavioral Sciences Team, the ‘US nudge unit’. Obama himself explained in a statement: “Adopting the insights of behavioural science will help bring our government into the 21st century in a wide range of ways – from delivering services more efficiently and effectively to accelerating the transition to a clean energy economy; to helping workers find better jobs, gain access to educational opportunity, and lead longer, healthier lives” (Office of the Press Secretary, 2015, p. 1). In Germany, the topic drew attention at the end of August 2014 when Chancellor Angela Merkel advertised three open behavioural insights positions to be housed within the Federal Chancellery. The public reaction was sceptical and some in the media painted the move in a negative light, with one headline in the German magazine *Die Welt* bluntly commenting: “Merkel wants to educate Germans with a nudge” (Dams, Ettl, Greive & Zschäpitz, 2015).

With more and more governments exploring ways to nudge citizens, opposition has swelled both in academic and public discourse. Many critics label nudging as a patronising and dubious form of control by which the state, lacking any democratic mandate, psychologically manipulates its citizens. Others argue that nudging exploits human weakness and shapes behaviour through opaque means, thereby raising complex ethical questions. Yet, common lines of criticism might not apply to all forms of nudging equally. Moreover, because nudges can be used to pursue different ends, depending on the goal or targeted behaviour, some nudges may be on more solid ethical ground than others. For instance, what about public policy aims that are supported by the majority of the population and important for the well-being of society, such as environmental protection? So-called *green nudges* function in the same way as nudges in other policy fields but specifically seek to encourage more sustainable and ecological lifestyles (Centre d’analyse stratégique, 2011, p. 1). Due to the dire, global, and even ‘super wicked’ nature of many environmental problems, is there a special argument to be made for green nudging (Levin, Cashore, Bernstein & Auld, 2012, p. 123)?

The present research centres on the question: Should we nudge for the environment? Much of the literature critical of nudge-based policies approaches the topic more broadly, examining all policy fields, but far fewer researchers assess nudging solely in the context of environmental policy. Nevertheless, nudging – whether for the environment or to pursue any policy aim – is an incredibly nuanced and complicated concept that deserves a careful and meticulous treatment. For one, nudging raises numerous ethical questions that either do not

¹ In 2014, the BIT became a social purpose company, but it is still partially owned by the Cabinet Office. For more information, visit: <http://www.behaviouralinsights.co.uk>

exist or are relatively straightforward for other policy instruments – e.g., taxes are a transparent means of behavioural control that do not diminish autonomy and personal agency, even though they can be inappropriately applied. The current paper serves first and foremost as a guide to help make sense of and assess all the existing positive and negative facets of green nudging systematically. Accordingly, this paper's primary objectives are (1) to synthesise and interpret the growing literature on green nudging, (2) present the current state of the art and existing practical applications of nudging and (3) establish a framework to evaluate the suitability of green nudging as an environmental policy instrument. In sum, with this paper we offer a synopsis of the most relevant and topical debates about green nudging as well as a rubric by which to evaluate this new policy tool.

The paper is structured as follows. In Section 2 we discuss the emergence of behavioural economics and how it serves as an empirical foundation for nudging. In this initial section, we also touch on the political philosophy of libertarian paternalism as well as the primary criticism and controversies surrounding nudge-based interventions. Section 3 turns the focus to green nudges. First, we examine the nature and complexity of environmental problems and the policy solutions traditionally used to take them on. We then place nudging in the context of the traditional environmental policy toolbox, highlighting how it is fundamentally different. In the latter half of Section 3 we propose a typology of green nudges and discuss the various categories using existing examples. Finally, at the end of Section 3 we revisit our research question against this theoretical and practical background.

In Section 4 we introduce the methodological approach used to develop a framework for evaluating green nudges. We then guide the reader through the proposed evaluative framework explaining the three chosen criteria: efficacy, effectiveness and ethics. We conclude this section with a discussion of methodological limitations. Section 5 serves as a preliminary evaluation of two types of green nudges – green defaults and green social norms. Each nudge type is assessed using the three criteria that comprise our framework. Finally, in Section 6 we discuss the outcome of the evaluation and embed our findings in the broader context of the current debate on green nudging and effective environmental policymaking. The discussion underscores the importance of evaluating nudging from an ethical angle and outlines recommendations for possible future research in the field. Section 7 concludes.

Compared to other policy instruments – such as effluent taxes and regulations – which have been around and in use for decades, nudging has only just recently emerged in environmental policy circles. Born of pioneering advances in the psychological and cognitive sciences, it is no wonder, however, that policymakers are just now beginning to discuss nudging as a possible alternative. Still, to adequately address fundamentally political questions about the suitability of green nudging, it is crucial to understand the scientific reasons why and how nudging changes human behaviour. The evaluative framework we propose takes these empirical foundations of nudging into account and is designed to be used by policymakers and other relevant stakeholders pursuing nudge-based interventions to pressing environmental

problems. It is our hope that by critically outlining the scientific foundations, practical applications and ethical considerations of green nudging as a new environmental policy tool, this paper will provide valuable input into the on-going debate.

2 Background: A theory of nudging

Nudge theory is founded on a rich and established body of empirical research in the behavioural sciences, notably experimental psychology and behavioural economics. This research has shown that small, unnoticeable changes to the context in which a decision is made can dramatically alter ensuing behaviour. Some of the more famous examples of nudges in practice include, the placement of a fly decal at the bottom of men's urinals to reduce cleaning costs, changing the renewable energy option default from 'opt-in' to 'opt-out' and the inclusion of a frowny face emoticon on electricity bills for households with exorbitant energy usage (Allcott & Mullainathan, 2010; Costa & Kahn, 2013; European Commission, 2016; Momen & Stoerk, 2014).² According to Thaler and Sunstein (2008), a nudge is defined as any element of the decision-making environment that influences choices without restricting options. Hence, as per their definition, the authors also introduce the topic with a political and philosophical twist, the crux of their thesis centring upon a freedom-preserving framework for public policy termed *libertarian paternalism*.

Libertarian paternalism, while seemingly an oxymoron, is any policy that upholds freedom of choice by unobtrusively guiding people towards a desired behaviour and thus can be considered a form of 'soft paternalism' (Sunstein, 2014, p. 20). In *Nudge: Improving decisions about health, wealth, and happiness*, the two authors claim that public policy-induced nudges do just this; they change behaviour without restricting or regulating options (Thaler & Sunstein, 2008). However, in the years since the 2008 publication, an expansive literature has emerged on the topic with many authors pointing out that nudging and libertarian paternalism, while related in theory, are not one and the same in practice (see for instance Barton & Grüne-Yanoff, 2015; Hansen, 2016; Hansen & Jespersen, 2013; Lepenies & Matecka, 2015; Mitchell, 2005). Indeed, Hansen and Jespersen (2013) discuss the differences between the two concepts at length, arriving at the conclusion that "while nudging is a means to promote behavioural change, libertarian paternalism is a guide, or a series of constraints on what ends may be promoted" (p. 12). As we will see, these and other critical treatments of nudging have helped reshape and add nuance to the discourse surrounding nudge theory.

A further distinction must be made between nudging as a policy tool to influence behaviour and so-called *behavioural insights* for policymaking more generally. Behavioural insights refer to behaviourally informed *input* into the policymaking process that helps tailor policies to the psychological realities of the human condition (European Commission, 2016, p. 10).

² Refer to Section 3.3 for a more exhaustive list of green nudge examples as well as a typology

Thus, all nudges can be considered behavioural insights, insofar as they are the result (or *output*) of behaviourally informed policymaking; but not all behavioural insights are nudges.

We contend that a comprehensive theory of nudging necessarily consists of two distinct concepts: (1) nudging as a psychological mechanism for guiding individuals' choices, informed by insights from the behavioural sciences and (2) the political philosophy of libertarian paternalism, which can – but need not always – serve as a guiding framework for the use of nudging in public policy. In our policy evaluation, we focus exclusively on the former, that is, green nudging as a tool or *means* to encourage sustainable and ecological behaviour, and we justify this approach in Section 3.4. However, the following will also outline in brief the tenets of libertarian paternalism.

This section is divided into two parts. First, we discuss the emergence of behavioural economics – the marriage of economics and experimental psychology – as a challenge to the classical economic model of rational human behaviour and how it serves as an empirical foundation for nudging. Second, we investigate the crucial differences between nudging and libertarian paternalism and explore how nudging can in theory be employed within a libertarian framework. Finally, we touch on some of the primary criticism of nudge theory and nudging in practice.

2.1 Nudging and its foundations in the behavioural sciences

Behavioural economics is a rapidly developing field at the nexus of cognitive psychology and economics, which upholds that a more comprehensive understanding of human cognition can inform, augment and enhance economic concepts (Thaler, 2000, p. 137). Since its inception in the mid-1900s, behavioural economics has come to play a prominent role in the policy-making of countries all over the world and, more importantly for our purposes, serves as the cornerstone of nudge theory.

2.1.1 Econs and Humans

For years, orthodox economic thinking has centred on a *theory of rational choice*, i.e., the notion that when it comes to making important as well as mundane decisions, humans are rational, calculative agents with stable preferences. Based on Daniel Bernoulli's essays on subjective utility in the 18th century and fine-tuned by John von Neumann and Oskar Morgenstern, into a series of axioms on *expected utility theory* in the mid-20th century, the assumption of rational choice informs the neo-classical economic analysis of policymaking – e.g., fiscal, monetary, health and environmental – to this day (Grüne-Yanoff, 2007, p. 353).³

³ Expected utility theory holds that you can measure an agent's preference for different *prospects* by the amount of risk he or she is willing to take on to obtain them. A *pure prospect* signifies zero risk or 100% certainty that the desired outcome will occur. For instance, the consumption of a good is a pure prospect (i.e. an apple will always provide a consumer with the associated amount of utility), while purchasing a lottery ticket is

Agents of the rational choice paradigm, sometimes called *homo economicus* or ‘Econs’, are omnipotent and consider all relevant information before making choices that inevitably promote self-interested ends (Thaler, 2000, p. 133; Thaler & Sunstein, 2008, p. 6). However, while the field of economics embraces the theory of rational choice or expected utility – both normative in nature – to explain how individuals should arrive at decisions, these rigid models are inconsistent with how the social and psychological sciences view human nature – i.e., positive or descriptive theories of decision making (Kahneman, 2012, p. 269; Thaler, 2000, p. 137).⁴ Behavioural scientists know from extensive research that humans are in fact ‘noisy’, unpredictable and (unlike machines) error prone even when it comes to their own best interests. Thaler and Sunstein (2003) go as far as to regard the historical assumption of perfect rationality in the field of economics as ‘tautological’. If relaxed to include psychological realities, many classical economic models would inevitably fall apart or cease to be valid (p. 176).

The rational choice model began to fray in the second half of the 20th century with the introduction of *bounded rationality*. The term was first coined by Herbert Simon in his 1957 book, *Models of Man*, but the author had already written extensively on the topic in earlier work, referring to ‘limited’ or ‘approximate’ rationality (see for instance Simon, 1955, p. 114). Unlike the prevailing theory, bounded rationality proposes that an individual’s ability to make rational decisions is subject to fundamental limitations, including *inter alia* the amount of information provided, the timeframe in which the decision must be met and internal cognitive restraints (Simon, 1972, p. 163). Expanding on this, recent research highlights the idea of *ecological rationality*, suggesting that individuals are confined by the environment in which a decision takes place (Gigerenzer, 2015, p. 374). It follows logically that this ecological/bounded notion of rationality is not a static concept and instead depends on context and circumstance. Therefore, while Econs operate under conditions of full rationality irrespective of context, ‘Humans’ must cope with existential and environmental bounds to their own decision-making faculties (Thaler & Sunstein, 2008, p. 6). Econs and Humans represent fundamentally different views on human psychology, with the latter reflecting reality and the former a more idealistic model of behaviour based on a theory of rational

not. The utility attached to the lottery ticket can be shown: $(X_1, p; X_2, p - 1)$, where X_1 is the utility derived from losing, p is the probability of losing, X_2 is the utility derived from winning (based on the magnitude of the prize), and $p - 1$ is the probability of winning. Agents choose prospects to maximize their expected lifetime utility. Expected utility theory has been the basis of economic thought since the end of World War II. (see Grüne-Yanoff, 2007, pp. 535-537 for a comprehensive explanation of expected utility theory).

⁴ By normative we mean that the theory of rational choice refers to the ‘ideal’ but non-existent consumer; it predicts how people *should* behave as opposed to positive theories, which describe how people *actually* behave. See Thaler (1980) for an overview of this line of reasoning.

choice with its own merits for stringent mathematical modelling (Thaler, 2000, p. 137). As we discuss later in Section 3.2, the degree to which either paradigm is taken as an analytical basis for policymaking has important implications for which policy tools are favoured.

Research in experimental psychology, such as the ground-breaking work by Nobel Laureate Daniel Kahneman and Amos Tversky, has identified a long list of ways in which humans fail to live up to the *homo economicus* or ‘Econ’ ideal. *Prospect theory*, for one, shows that humans are both *loss averse* – i.e., we dislike losses more than we like gains of the same magnitude – and *risk averse* – i.e., when it comes to gains we exhibit a preference for certainty over gambles, even when the gamble has a higher expected utility – (Kahneman & Tversky, 1979, p. 280, 1984, p. 342). Conversely, humans are *risk seeking* for potential losses. Unlike expected utility theory, prospect theory also demonstrates that individuals measure changes in wealth based on the status quo rather than end quantities of wealth. Consequently, humans are biased in their subjective reading of utility to attach undue weight to anything already in their possession – the so-called *endowment effect* (Thaler, 1980; Tversky & Kahneman, 1991).

Further research suggests that people think of their money in an organized way; that is, consumers budget planned expenses and attach emotions to these *mental accounts* (Thaler, 1985). The concept of mental accounting explains why ‘sunk costs’ weigh so heavily in human decision making. For example, we are more likely to continue investing in a project that has already wasted a considerable amount of time and money than in a brand new one with the same chance of success because, unlike Econs, Humans are swayed by the need to justify past expenses (Kahneman, 2012, p. 345).

The above examples represent a snapshot of the cognitive tendencies that underlie the field of behavioural economics and throw into relief the practical and theoretical limitations of expected utility theory. For their contributions, psychologists, Kahneman and Tversky, along with the economist, Richard Thaler, are generally credited as the founders of behavioural economics and by extension its application to policymaking (i.e. nudging). Nudge theory emphasizes primarily the internal, cognitive limitations common among all people (Schubert, 2016, p. 3). Indeed, nudging is more than simply influencing behaviour without limiting options and may be more readily defined as the exploitation of bounded human rationality by either public officials or private enterprises (i.e. marketing and advertising) with the goal of influencing behaviour. Put slightly differently, a nudge is “any factor that alters the behaviour of ‘Humans’ (irrational, cognitively biased actors), even though it would be ignored by ‘Econs’ (rational actors)” (Thaler & Sunstein, 2008, p. 8). To fully grasp the built-in limits to human judgment and how nudge theory in turn makes use of bounded rationality to induce behavioural change, we must first take a look at the prevailing model of human cognition.

2.1.2 Dual process theory

We spend our days transitioning from one choice to the next. Consciously or not, the human brain is perpetually in the business of making decisions. The principle set of theories underlying much of our current understanding of human cognition and decision making is known in conjunction as *dual process theory* (Evans, 2008, p. 255). Dual process theory describes two processes that interact to instruct the way we evaluate and select among options: an automatic, intuitive process and a reflective, deliberate process – hereafter referred to as System 1 and System 2, respectively (Evans & Stanovich, 2013, p. 244). Generally, System 1 operates involuntarily and quickly, while System 2 consciously allocates mental effort and attention to carefully evaluate alternatives. It may be helpful to think of System 1 as impulsive and System 2 as just the opposite, cautious and planned (Kahneman, 2003, p. 698).

A slightly different but related thesis is the *Elaboration Likelihood Model* (ELM) for attitude adjustment and persuasion, which consists of the peripheral and central routes of information processing (Petty & Cacioppo, 1986).⁵ Similar to System 1, the peripheral route picks up on and responds to external cues of little personal interest when evaluating information, while the central route thoughtfully considers personally relevant alternatives, thereby appealing to System 2. Hence, the ELM arrives at many of the same conclusions as dual process theory, but from a slightly different angle.

Generally speaking, System 1 and System 2 operate in concert, with conclusions drawn intuitively by System 1 often informing the calculative, attentive nature of System 2 (Hansen & Jespersen, 2013; Michalek, Meran, Schwarze & Yildiz, 2015). In other words, System 2 is there “to monitor and control thoughts and actions ‘suggested’ by System 1, allowing some to be expressed directly in behaviour and suppressing or modifying others” (Kahneman, 2012, p. 44). Thus, whenever an individual makes a decision – for instance, whether to select the renewable energy option from their local utility – both systems may engage (Momsen & Stoerk, 2014, p. 376). However, System 2 processes are slower to operate or may never kick in at all, thereby leaving many simple, everyday tasks to System 1. For this reason, in his book *Thinking Fast and Slow*, Daniel Kahneman refers to System 2 as “the lazy controller”, noting that volitional and analytical thinking requires mental effort (p. 39). In the renewable energy example, System 2 may not engage quickly enough to weigh and judge all available options – i.e., ‘yes/no, I want to source 50% of my electricity sustainably’ (Momsen & Stoerk, 2014, p. 378). The decision is then left to System 1, which in this case is predisposed to stick with the status quo or whichever option is presented as the default. This *status quo bias* is

⁵ Petty and Cacioppo (1986) showed in their research that information provision targeted at the central route of cognition (designed to appeal to personally-relevant internal and reflective processes) tended to have a longer lasting influence on attitudes and resulting behaviour. This may have implications for nudge theory. Nudges that function solely by influencing automatic behaviour – i.e., habit manipulation or what Hansen and Jespersen (2013) refer to as ‘Type 1’ nudges – may not have a lasting effect on behaviour.

one of many cognitive biases that System 1 falls prey to (we discuss biases at length in the following section).

The dual processes of cognition are at the centre of nudge theory. Nudges as policy tools function first and foremost at System 1 levels of cognition, influencing automatic and involuntary behaviour and decision making (Hansen, 2016, p. 8; see also Grüne-Yanoff & Hertwig, 2016). For instance, in the example above, nudging people by changing the default to the more sustainable option, has been shown to lead to a 20% increase in the purchase of renewable energy by households in experimental settings (Momsen & Stoerk, 2014, p. 380). Even though System 2 may remain latent at first, once engaged, it interacts with System 1, choosing either to ignore or use cognitive input from the automatic process. Thus, depending on how a nudge is designed, it can also indirectly influence the reflective and self-aware decision making of System 2 (Hansen & Jespersen, 2013, p. 15).

A treatment of nudging that omits the basic premises of dual process theory leads to a “blurring” in the definition of what constitutes a nudge (Michalek et al., 2015, p. 5). This may explain some of the inconsistency in the literature, which at times, struggles to pin down a single psychological definition of nudging (Hansen, 2016, p. 2). The current paper uses a definition of nudging based on the premises of dual process theory and assumes the position furthered *inter alia* by Hansen and Jespersen (2013), Schubert (2016) and Sunstein (2014) that nudges operate primarily via System 1 biases and heuristics but, due to interaction between the two systems, have the ability to affect more deliberate decision making.

2.1.3 Heuristics and biases

Sometimes System 1 becomes overloaded and relies on *heuristics* to cut down on the effort required to undergo a particular task. The psychological definition of a heuristic is any “simple [cognitive] procedure that helps find adequate, though often imperfect, answers to difficult questions” (Kahneman, 2012, p. 98). In other words, heuristics are mental shortcuts that function through so-called ‘attribute substitution’, which basically entails swapping a hard problem for an easier one (Kahneman, 2012, p. 97; Michalek et al., 2015, p. 5). For example, if asked to estimate the probability of a given event occurring, few people will employ System 2 to search out the needed information and complete the necessary calculations. Instead, most will judge *something else* (not probability) and still provide an answer in probabilistic or quantitative terms. Kahneman (2012) points out that in the majority of cases willingness to pay paradigms used to appraise, e.g., environmental issues exemplify a heuristic at work (p. 98). If a person is asked how much he or she is willing to pay to save a neighbourhood green space from development (hard question), they may unconsciously consider the emotions they associate with the place (easy question). It is plain to see that System 2 must not engage to provide an adequate answer, even in monetary terms. A high degree of positively valenced emotion in this instance will inevitably lead to a higher willingness to pay. This *affect heuristic* describes how judgments are made based on the emotions attached to alternatives (Slovic, Finucane, Peters & MacGregor, 2002). Perceived emotion associated with a given choice serves as a substitute for cost-benefit or quantitative

analysis and prompts a situation in which the person's "likes or dislikes determine their beliefs about the world" (Kahneman, 2012, p. 103).

In addition to the affect heuristic there are three principle heuristics that affect human judgement: availability, representativeness and anchoring (Tversky & Kahneman, 1974).

- The *availability heuristic* suggests that when making a judgment, people generally rely on that which easily comes to mind either because of personal or temporal salience. For example, the decision to buy fire insurance (a probabilistic judgment) is influenced by the experience of witnessing a neighbour's house on fire or watching a house fire on the news.
- The *representativeness heuristic* occurs when an evaluation is made based on how similar two events or instances seem as well as whether one belongs to or originates from the other. A clear example here is the propensity, even among statisticians, to consider correlation as evidence of causation. Just because two phenomena are on similar trajectories does not imply they are intrinsically related, much less causally related.
- The *anchoring (and adjustment) heuristic* explains how judgments are made by beginning with an initial value and then adjusting to obtain a final answer. For instance, if asked to estimate the population of Essen (a medium sized city in central Germany), an individual may begin with a known value (e.g. the population of Berlin) and adjust downwards to come up with an approximation.

Similar to how at first promising shortcuts can lead one astray while out for a walk in unknown territory, these and other heuristics often lead to *biases* or systematic flaws in our thinking (Ölander & Thøgersen, 2014, p. 344; Tversky & Kahneman, 1974, p. 1124). The availability heuristic, for instance, explains a bias towards *cognitive ease*, i.e., that which can be recalled or processed easily is more likely to be factored into decision making. Kahneman (2012) refers to this phenomenon broadly as 'WYSIATI' or 'what you see is all there is', claiming that it serves as the basis for three additional biases: (1) *overconfidence* in the quality of judgments, (2) *base-rate neglect* (i.e. failing to abide by Bayesian reasoning) and (3) *framing effects* (p. 82).⁶

⁶ Thomas Bayes' rule of probability demonstrates how knowledge of base rate statistics (i.e. fundamental distributions within a population) can be combined with other more specific statistical observations about the population to arrive at a more accurate estimate of an occurrence or phenomenon. For example, to estimate the probability that a given individual belongs to a specific group (e.g. engineering students at University X who are female), you must first account for the base rate (e.g. 1 in 10 students studies engineering at University X). Only then you can take other statistical information into account (e.g. there are 2x more male than female engineers). (See Kahneman, 2012, p. 154 for a more in-depth explanation of Bayesian probability.)

Framing effects are caused by the way in which available information is presented, which at times can lead to less-than-rational decisions. One famous example is the Asian Disease Paradigm described by Kahneman and Tversky (1984). In two separate instances, the researchers asked participants to choose between programmes to combat a hypothetical deadly outbreak in Asia. For the first choice, Programme A ensured that 200 people would be saved and Programme B entailed a one-third chance that 600 people would be saved and a two-thirds chance that no one would be saved. A clear majority of participants chose Programme A, suggesting that by avoiding the gamble the sample was risk averse. The participants were then asked to choose between two additional alternatives: Programme C, which allowed 400 people to die, and Programme D, which entailed a one-third chance that nobody would die and a two thirds chance that 600 people would die. Interestingly, this time participants chose the latter option. Upon closer look, it becomes obvious that the two choice pairs are equivalent. Yet, the decision between Programme A and B is framed in terms of ‘lives saved’, while the decision between Programme C and D is framed in terms of ‘lives lost’ – in other words, gains and losses – which, due to loss aversion, significantly influenced how individuals judged the decision (Kahneman & Tversky, 1984, pp. 343-344).

In the prior example the framing effect originated from the way in which the decision was worded, but other attributes can be altered to produce similar results. Both attribute salience and simplification impart additional information or cause mental associations that are picked up primarily by System 1 processes, in effect changing the preferences of the decision maker (Michalek et al., 2015, p. 9). For instance, an ever-growing body of psychological research has demonstrated that social norm salience can induce a significant framing effect (see for instance Cialdini, 2005; Cialdini, Reno & Kallgren, 1990; Costa & Kahn, 2013; Graffeo, Ritov, Bonini & Hadjichristidis, 2015). Social norms are different than personal norms or values because they evoke associations of belonging, conformity and societal sanctions, and thus have a strong and stable influence on human behaviour (Kinzig et al., 2013, p. 165). Decisions can be framed using two distinct normative categories: descriptive (i.e. information on what is generally done) or injunctive (i.e. information on what society deems morally right or wrong) (Schultz, Nolan, Cialdini, Goldstein & Griskevicius, 2007, pp. 429-430).⁷ Research has shown that framing effects are influenced by whichever social norm is more salient, a phenomenon called the *focus theory of normative conduct* (Cialdini et al., 1990). In other words, descriptive and injunctive normative framings can produce independent effects on human judgement, even if presented simultaneously (Schultz et al., 2007).

Default options, such as the renewable energy example mentioned above, change the format in which information is presented and thus can also be considered a form of framing. The

⁷ We include the third category, *comparative*, in our typology in Section 3.3 to denote when a social norm is presented in such a way that makes a comparison of behaviour more salient.

status quo bias that pins individuals to a default option is fundamentally caused by a combination of the anchoring heuristic and loss aversion, but defaults are also thought to function via *cognitive inertia*, which is a fundamental aversion to mental strain (i.e. a bias for cognitive ease) as well as *implicit recommendation*, which implies that one perceives default options as being superior or endorsed (Kahneman & Tversky, 1984, p. 348; Michalek et al., 2015, p. 9).

Additional biases often described in the literature include *present bias*, *confirmation bias* and the *overthinking bias*. Present bias concerns our propensity to discount gains and losses that we foresee in the future (European Commission, 2016, p. 8). Although discounting is a fundamental aspect of economic analysis, such as cost-benefit analysis or cost-effectiveness analysis, both the anchoring and availability heuristics help explain this bias for the present from the perspective of cognitive psychology. Moreover, present bias can also be thought of in terms of time-inconsistent or *hyperbolic discounting*, which refers to experimental evidence that people disproportionately discount distant-future prospects compared to near-future ones (Thaler, 2016, p. 91). Confirmation bias occurs due to associative memory and the availability heuristic (or that which comes to mind easiest), and can be described as the “deliberate search for confirming evidence” (Kahneman, 2012, p. 81). Confirmation bias may help explain instances of science denial; those who are sceptical of anthropogenic climate change (which enjoys a high level of scientific consensus), may be trapped by a System 1 bias for confirmatory evidence found outside of scientific circles that vindicates their own pre-existing beliefs (Weber & Stern, 2011, p. 5). Finally, the overthinking bias is actually a System 2 bias that is caused by increased levels of so-called ‘cognitive noise’ or information and stimulant overload that leads to mental strain and subsequently indifference or preference inconsistency when faced with a decision (Lee, Amir & Ariely, 2009, p. 173). Moreover, after long-term cognitive strain, an individual may inadvertently revert to relying on System 1 processes, thereby perpetuating all the aforementioned biases for tasks that require more deliberate thinking.

Together the findings from cognitive psychology detailed above constitute the *heuristics and biases approach* to bounded rationality. With this rich body of research as an empirical basis, in the next section, we explain how nudging utilizes the heuristics and biases associated with System 1 cognition to change behaviour and the choices individuals make.

2.1.4 Designing a nudge: Choice architecture

Thaler and Sunstein (2008) aptly use the phrase *choice architecture* to signify the context in which a decision is made, implying that like any physical structure, the environment surrounding a decision can be ‘designed’ by ‘choice architects’ (p. 3). Discussed at length above, framing alters the choice architecture, guiding the decision maker predictably in one direction or another. Consequently, nudging can be thought of as the intentional framing of a choice by a political or private actor. Indeed, most nudges in use today (e.g. defaults, social norms, feature salience and others) operate chiefly via framing effects *in addition* to a handful of other biases. *Priming*, however, is one example where this is not the case.

Priming nudges expose the individual to some stimulus *before* any choice is presented. The primed stimulus is held unconsciously in associative memory and influences subsequent decisions and behaviour due to the availability heuristic (Momsen & Stoerk, 2014, p. 379). Not to be confused with subliminal messaging, which makes use of covert, rapidly presented images or text; the prime itself is not necessarily unbeknownst to the individual even though its effect occurs subconsciously. In Section 3.3 we describe the different types of nudges that exist in the environmental field and propose a typology that organizes these nudges based on their function.

Recall the definition of a nudge proposed at the onset of this paper as any aspect of the decision context or choice architecture that will lead to a predictable change in behaviour without limiting the chooser's options (Thaler & Sunstein, 2008, p. 6). It may now be clear to the reader that this definition is far too simple. A more exact definition would necessarily take both dual process theory and the heuristics and biases approach to bounded rationality into account. Thus, in Figure 1 we propose multiple revisions to Thaler and Sunstein's frequently cited original description and arrive upon a 'technical definition' of a nudge, which we then use throughout the remainder of this paper.

A *nudge* is any part of the choice architecture that systematically capitalizes on System 1 biases to guide decisions and/or behaviour in a predetermined and predictable way; it may but must not necessarily uphold freedom of choice.

Figure 1: A technical definition of nudge

Nudges function first and foremost by either capturing and utilizing or blocking the biases that arise from System 1 thinking. We include the word 'predetermined', for the same reason Hausman and Welch (2010) refer to nudges as "ways of influencing choice", in order to connote agency (p. 126). Thaler and Sunstein (2008) claim that nudges are not always intentionally placed, and even go a step further arguing famously that there is no such thing as a 'neutral design' for choice architecture – in other words, we are constantly being nudged whether there is intention behind the nudge or not (p. 11). However, in this paper we only consider nudges that are intentional, as assuming agency on behalf of the policymaker is a precondition for evaluating nudging as a policy instrument. A similar approach is taken by others (e.g. Lepenies & Małecka, 2015).

The second part of our definition pertains to whether the nudge approach restricts choices. From our perspective (one grounded in a behavioural science understanding of nudging), upholding freedom of choice (i.e. libertarian paternalism) is not required for something to qualify as a nudge. This runs counter to how many authors define nudging (including Thaler and Sunstein), who arrive at their definition through a political, philosophical lens. That said, in the following we discuss nudging as a *policy* to induce ecologically friendly behaviour. As much of the literature on nudging as public policy centres on the political paradigm of libertarian paternalism, we consider it necessary to subsequently discuss this in brief. The

concept of libertarian paternalism has implications especially for ethical treatments of nudging and in recent years has become a hotly debated topic among politicians, legal scholars, philosophers as well as behavioural scientists.

2.2 Nudging and libertarian paternalism

A policy or government is paternalistic if it limits individual liberties by banning, regulating or influencing behaviour or choices to achieve the governing authority's own aims. Such a policy may also aim to increase social welfare more generally but does so by targeting the choices made by individual citizens. The notion of government promoting or prohibiting certain behaviours has been the topic of endless debate throughout history. Drawing from John Stuart Mill's famous critique of paternalism in *On Liberty* (1863), in which he denounces "despotic, or what is called paternal, government", ethical concerns regarding paternalism – including most current arguments – centre on the infringement of rights and freedoms (p. 197). Mill sums up his condemnation of paternalistic society, claiming: "The sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection" (p. 23). Sunstein (2014) and others refer to this principle of the rights of government as Mill's *Harm Principle*, and generally use it to frame discussions of potentially paternalistic policymaking (p. 5). Simply put, the normative argument follows that policymakers should refrain from intruding into the lives of citizens unless they seek to save some number of them from harm.

Legislation that aims at changing behaviour in the absence of harm to others is paternalistic; however, the degree of paternalism can be viewed as existing on a spectrum (Sunstein, 2014, pp. 55-65). On one end, *hard paternalism* imposes high costs on individuals, material or otherwise. Hard paternalism examples include most regulatory policy interventions that induce behavioural change by threatening certain sanctions. A sin tax on cigarettes, for instance, could be considered a case of hard paternalism, as those affected must choose to either give up a portion of their income or cease the behaviour. Conversely, *soft paternalism* entails few or negligible costs for individuals. To stick with the example of financial costs, a minor five-cent charge for grocery bags is categorized more readily as soft paternalism. Under this framework, libertarian paternalism can be considered a special form of soft paternalism that does not involve any *material* costs (Sunstein, 2014, p. 57). Put differently, libertarian paternalists argue that a policy intervention can be designed in such a way that guides behaviour while upholding an individual's freedom of choice, in effect circumnavigating the most common objections to paternalistic policymaking. Thaler and Sunstein (2003) claim that even though "libertarians embrace freedom of choice [...] and paternalists are thought to be deeply sceptical of freedom of choice", libertarian paternalism should be viewed positively as a promising way to bridge gaps between polarized political ideologies (p. 16).

Proponents of nudging (most notably Thaler and Sunstein) consider the nudge approach as inherently libertarian, and use this to justify nudging as an acceptable flavour of soft paternalism. Moreover, Sunstein (2014) argues that due to the prevalence of 'behavioural market

failures' (i.e. the heuristics and biases outlined above), the notion that individuals know what is in their best interest can be contested (p. 7). As a consequence, there is a clear *moral* argument, akin to Mill's *Harm Principle*, for using libertarian paternalistic interventions such as 'pro-social' nudges to steer people away from decisions that harm society and 'pro-self' nudges, which aim at protecting the individual from bad decisions (Barton & Grüne-Yanoff, 2015, p. 344). Nonetheless, in a separate book entitled *Why Nudge?*, Sunstein concedes that not all costs related to policy intervention are material or financial and that nudging may require a temporary decrease in well-being or impose unavoidable psychological costs, such as disgust towards the packaging of cigarettes, which in some countries contains graphic or disturbing imagery.

Paternalistic interventions can be further categorized as ends-based or means-based (Sunstein, 2014, p. 63). *Means paternalism* responds to and targets people's own ends, giving them the autonomy to choose whatever these may be. Therefore, means-based nudging attempts to guide an individual towards behaviour that is in their own interest, "as judged by themselves" (Thaler & Sunstein, 2008, p. 5). On the other hand, *ends paternalism* steers people towards set ends regardless of whether these are wanted or not. Most pro-social nudges, fall in this category because the specific end of the nudge may or may not align with the personal ends of the individual being nudged. The renewable energy default nudge, for example, is arguably good for society – the uptake of solar and wind technologies reduces the burning of fossil fuels, which release climate change-causing greenhouse gasses (GHG) and other harmful pollutants. Yet, not every person sees clean energy use as a personal end. Thus, Thaler and Sunstein's concept of libertarian paternalism as chiefly means-based may be too narrow because the degree to which individuals can and do decide their own ends under a nudge-based regime is questionable.

Some claim that nudging and libertarian paternalism are often conflated and accordingly that libertarian paternalism alone is an unsatisfactory justification for policy-induced nudges (see for instance Gigerenzer, 2015). Hausman and Welch (2010) contend that many of the examples called nudges and thus by extension labelled as libertarian paternalism are not paternalistic at all. For example, information campaigns that shed light on the harms of smoking or use descriptive social norms to reduce binge drinking behaviour on college campuses are not paternalistic, as their primary aim is to protect a population from harm, while concurrently respecting rights and treating individuals as autonomous decision makers (Hausman & Welch, 2010, p. 127).⁸ In brief, there seem to be two camps in the literature: those

⁸ As per the definition of nudge in Figure 1, the anti-smoking campaign alone does not qualify as a nudge because it appeals primarily to System 2, reflective processes. Nevertheless, many authors describe pure information provision as nudging, which we consider to be a major cause for confusion and inconsistency within the literature. Certainly, de-

who claim that all nudges are a form of libertarian paternalism and defensible as such (e.g. Thaler & Sunstein, 2008) and those who argue that nudges are conceptually separate and only in some cases follow the tenets of libertarian paternalism (e.g. the current authors; Hansen, 2016; Hansen & Jespersen, 2013; Hausman & Welch, 2010). The remainder of this document concerns an evaluation of nudging as an environmental policy tool and does not investigate libertarian paternalism as such. However, particularly in our assessment of the ethics of nudging, the discussion returns to issues related to autonomy and human dignity. It is therefore necessary to keep the political philosophy of libertarian paternalism in mind.

2.3 Criticism

In the decade since its introduction, nudge theory has attracted a great deal of criticism regarding its effectiveness, ethical grounding and broader role in public policy. From a practical standpoint, some question the real-life usefulness of a policy informed by bounded rationality. For one, Gigerenzer (2015) claims that nudging puts too much emphasis on the ‘individual mind’, in effect “closing our eyes to institutions that steer behaviour so that they can take advantage of it” (p. 363). In other words, nudging does not adequately take ecological rationality into account and is instead founded on ‘narrow logical norms’ of human rationality (Gigerenzer, 2015, p. 365). In a sense, the author’s argument boils down to the idea that nudging underestimates human capacity to make informed decisions and adapt behaviour. This argument is echoed by Lepenies and Matecka (2015), who suggest that a ‘non-cognitive’, nudge-based policy undermines individuals’ ability to ‘self-legislate’, i.e., learn, adapt and change behaviour wilfully (p. 435). By way of a metaphor, many academics view nudging as a form of government induced autopilot. Similar to how a pilot can use autopilot to guide his or her aircraft, thereby avoiding mistakes that could lead to disastrous results, governments can rely on System 1 biases to steer citizens in one way or another. However, neither the pilot nor the citizen learns anything from this setup and once autopilot is removed both must return to their biased and faulty decision-making facilities. This relates to a further common criticism, namely that there is little evidence to suggest that the behavioural change caused by a nudge is long-lasting (Raihani, 2013, p. 2). Especially, nudges that work solely at System 1 levels of cognition, impacting only automatic, habitual processes, may lose their effectiveness over time (Michalek et al., 2015, p. 14). Furthermore, studies show that social norm nudges are mediated by political ideology, suggesting that individual difference variables such as *inter alia* culture and political persuasion limit the scalability of some nudges (Costa & Kahn, 2013).

pending on how the campaigns are designed and implemented, nudges could be incorporated by, e.g., framing the information in a particular way or evoking social norms (such as in the binge drinking example).

From an ethical standpoint most detractors paint nudging as a ‘manipulation of choice’ (Bovens, 2009, p. 208). Common arguments against nudging claim that it is ‘non-transparent’, operates ‘in the dark’ and infringes upon the personal autonomy by acting on behalf of individuals without their explicit permission (Bovens, 2009, p. 4; Hansen & Jespersen, 2013, p. 15; see also Schubert, 2015, 2016; Sunstein, 2015). The so-called *welfarist argument* sums up these and other moral arguments against nudging, suggesting that due to the inevitable infringement of human autonomy and dignity, subjective well-being is in turn affected (Sunstein, 2015, p. 414). In other words, robbing individuals of their agency in a way that is non-avoidable and not transparent will lead to a decrease in welfare because autonomy is a right. Furthermore, Hansen and Jespersen (2013) and others bring up the crucial matter of accountability on the part of the one doing the nudging, i.e., the government. If used malevolently by a repressive or dictatorial regime, then nudging is clearly on weak ethical grounds (Hansen & Jespersen, 2013, p. 4). Finally, Lepenies and Matecka (2015) advocate for restraint in the use of nudging, highlighting the legal and institutional ramifications of injecting nudging into policy. The authors’ central thesis contends that nudges are ‘non-normative’ policy instruments, which influence behaviour without using legal norms. Furthermore, because nudges are invisible to the public they are ‘diachronic’ in nature, meaning that they may automatically persist whenever the government changes hands among political parties and in the worst case may simply be forgotten by public officials (Lepenies & Matecka, 2015, p. 435).

Above we detail some of the many open questions concerning the use of nudging as public policy. Due to the complexity of the field and fierce debate surrounding it, incorporating nudge theory into policymaking is no straightforward task (Moseley & Stoker, 2013). Yet, as we will see, particularly in the realm of environmental policy, there may be a role for nudging, if implemented in such a way that respects the overall ethical and practical limitations of the tool. In the next section, we turn specifically to ecological or *green nudging* and discuss the ways in which nudge theory can and has been used to pursue environmental ends.

3 Nudging for the environment

So far we have outlined the origins and functionality of nudging more broadly. In the following, we focus in particular on nudging in the environmental field, what we call green nudging. First, this section will outline the nature and complexity of environmental problems. Next, we will discuss the policy instruments that have been used thus far to tackle environmental problems and how successful they have been. Afterwards, nudging will be introduced as a possible addition to the environmental policy toolkit, better suited to address the bounded rationality of human nature than the conventional instruments. Additionally, the characteristics and appropriateness of nudges in the environmental field will be laid forth. In the subsequent section, a typology of green nudges is presented to allow for a more nuanced view of nudging as an instrument with different dispositions and effects. Lastly, against this background we pose the guiding question of this paper, namely whether nudges should be used as a means to pursue environmental ends.

3.1 The nature of environmental problems

By virtue of the focus of this study, we only address problems which emerge from “environmentally significant” human behaviour and not natural disasters or chance occurrences (Stern, 2011, p. 305). Many of the negative impacts of human consumption and activity on the environment have certain characteristics that distinguish them from other policy areas and make them especially difficult with respect to traditional policy instruments. Such features are, for example, that they are usually complex, concern long time horizons and involve consequences and causes that are often unequally distributed (Mickwitz, 2003, p. 417). Furthermore, the environmental protection is a public good, for example, clean air, fresh drinking water, arable soil etc. For this reason, socio-political factors such as social (in)equality, economic and political participation rights as well as social inclusion are highly relevant regarding the provision of environmental goods (Croson & Treich, 2014, p. 336).

3.1.1 The tragedy of the commons

The lack of noticeable cause-consequence relations in terms of the treatment of environmental goods as well as their public character often lead to environmental devastation. The concept of the *tragedy of the commons*, most famously posited by Garret Hardin in the 1960s, explains this phenomenon. The theory basically states that individuals put their own self-interest over the interest of the community when it comes to the usage of public resources (Hardin, 1968). This happens primarily because commons resources are not anyone’s private property. Consequently, no one feels personally responsible for the good/resource and no one can be excluded from consumption (e.g. breathing fresh air). Moreover, the negative impacts of private actions may not have a direct or immediate effect on the polluter and are instead diffused in society (Ostrom, 1990). The fact that a single individual can benefit from common goods and natural resources without the risk of being excluded from their use if he or she does not contribute to its sustenance, opens the door to ‘free-riding’. The fundamental problem with most natural resources is that the incentive for the individual to ensure that the public good is preserved is simply not large enough (Grossman & Hart, 1980, p. 42). Additionally, this refusal to contribute usually remains unnoticed by the rest of the population, since each person’s share is not large enough to make a direct, visible difference in outcome and thus cannot be traced back to the responsible polluter or consumer (Grossman & Hart, 1980).

One good example of both a tragedy of the commons and the free-riding dilemma is human-induced climate change (Dietz, Ostrom & Stern, 2003, p. 1907). In this case, the negative consequences of climate change caused in part by individual or company behaviour (e.g. flying on a regular basis or producing energy-intensive products) do not directly impact the polluter. Generally, the consequences become noticeable in other parts of the world where the conditions are already fragile (e.g. small island countries) or where people are in more vulnerable climatic positions (e.g. those living under conditions of drought in Africa). Furthermore, these consequences may only become apparent in the distant future (e.g. the United Nations Development Programme (UNDP) predicts that by 2050 200 million refugees

will be forced to leave their country due to climate change and need support from developed nations) (Brown, 2007, p. 2). Possible consequences of climate change induced mass migration include the depletion of public resources and conflict (Dietz et al., 2003).

Interestingly, the original game theoretical formulation of the tragedy of the commons assumes purely rational, self-interested agents and does not take behavioural considerations into account. However, extensive experimental research conducted by Elinor Ostrom and her colleagues has underlined the limits of this *homo economicus* approach to conceptualising social dilemmas. She argues persuasively that a rational choice approach to commons and social dilemmas views the individuals involved as “prisoners” unable to escape the tragic conflict between private and social interest (Ostrom, 1990, p. 7). However, laboratory and field research suggests that in reality a myriad of behavioural factors, including the powerful influence of social norms, work to both facilitate and hinder collective action (Ostrom, 1998, 2003a).

3.1.2 Barriers to ecological behaviour

As we have seen, the nature of environmental problems makes them especially difficult to solve. Apart from the general problems related to common resources mentioned above, other psychological, financial and material obstacles stand in the way of adopting ecological behaviour (Centre d’analyse strategique, 2011, p. 1).

Looking at households, consumption at this level is accountable for a relatively large share of global GHG emissions – either directly through the use of products and services or indirectly through domestic as well as foreign supply chains (Michalek et al., 2015, p. 1). Thus, fostering environmental friendly behaviour at the household level can contribute considerably to reduced GHG emissions and potentially mitigate climate change (Michalek et al., 2015). However, to do so, individuals must be addressed directly and their daily habits altered. This is where psychological factors such as bounded rationality become apparent, which can be difficult to address with public policies.

There are five elements specific to environmental issues that, due to human psychology and bounded rationality, hinder ecological behaviour. These include passive choice, complexity, limited personal experience, third-party marketing and inter-temporal choice (Beshears, Choi, Laibson & Madrian, 2008, pp. 1788-1790). In addition, environmental issues are often associated with strong moral feelings such as guilt, shame, pride, etc. These feelings affect people’s beliefs and attitudes towards environmental programmes and policies (Croson & Treich, 2014, p. 336). Therefore, when policymakers try to push people too hard into a certain type of desired behaviour, these efforts tend to fail due to psychological barriers. One other reason for this is that when it comes to environmental policies, the factors of complexity, long time frames and vague cause-consequence relations result in political measures in which the direct negative outcomes for the individuals (e.g. restrictions or payments) are much more noticeable than the positive outcomes that will materialize for society in the long run. Since payoffs or damages do not always become immediately apparent,

there is no perceived need for fierce action – e.g., in the case of climate change, the benefits of mitigation via emission reductions are spread temporally but the costs are borne in the present (Centre d’analyse strategique, 2011, p. 3).

There are also material factors that reinforce the barriers to ecological behaviour. Many environmental friendly habits can only take place if the necessary infrastructure is provided (e.g. recycling system, public transportation system, etc.). Furthermore, these services must be accessible for everyone. The less developed such structures are, the less likely they will be used (Kollmuss & Agyeman, 2002, p. 248).

Finally, the so-called attitude-behaviour gap plays an important role when motivations for or against ecological behaviour are being discussed.⁹ The concept describes an observed discrepancy between knowing about or being willing to change environmentally harmful behaviour and acting to do so. Ecological attitudes tend not to be reflected by corresponding forms of behaviour if active changes in daily routines and habits are required (Centre d’analyse strategique, 2011, p. 2). Even if knowledge is acquired, it does not necessarily cause the individual to abandon the environmentally destructive habits (Centre d’analyse strategique, 2011, p. 2). Psychologists as well as social and political scientists have tried to find ways to close the attitude-behaviour gap with a range of different soft policy tools such as information and market-based incentives (Kollmuss & Agyeman, 2002).

Until recently, most theories in environmental economics based their argumentation on a neoclassical model, assuming humans to be rational actors or *homo economicus*. Consequently, conventional recommendations focus on policy instruments that rely on sanctions, incentives or the provision of information (Schubert, 2016, p. 2). As described in the previous section, there has been a shift in recent years towards a more realistic picture of human economic behaviour. This shift has led to the incorporation of behavioural insights in the field of environmental policy. Among such strategies, green nudges may become a popular means of addressing environmental problems. The aim of a green nudge is to use “behavioural biases such as compliance to social norms or inertia to change in order to encourage citizens to adopt lifestyles showing a greater respect for the environment” (Centre d’analyse strategique, 2011, p. 1). In the next section, we introduce ways in which green nudges can complement the traditional environmental toolkit and outline what sets them apart.

3.2 Nudging as an addition to the traditional policy toolkit

Green nudges expand the range of policy instruments used to tackle pressing environmental concerns. Below we outline the commonly used or *conventional* environmental policy instruments. Thereupon an attempt will be made to incorporate green nudging into this scheme as a potential addition to the traditional policy toolkit. In this process, the characteristics

⁹ The attitude-behaviour gap is also commonly referred to as the *intention-action* gap (e.g. Allcott & Mullainathan, 2010; Momsen & Stoerk, 2014).

of nudging as a new non-regulatory, soft policy tool will be examined, different types will be presented in detail and the adequacy of green nudges will be questioned.

3.2.1 Traditional environmental policy instruments

Environmental policy instruments as defined by Mickwitz (2003) are “the set of techniques by which governmental authorities wield their power in attempting to affect society – in terms of values and beliefs, action and organization – in such a way as to improve, or to prevent the deterioration of, the quality of the natural environment” (p. 419; see also Lundqvist, 1996, p. 16; Vedung, 1998, p. 21).

Table 1: Traditional environmental policy toolkit

RESTRICTIVE <<<		>>>NON-RESTRICTIVE				
Type	Regulatory		Economic		Information	
Intervention category	Eliminate choice	Restrict choice	Market-based	Incentive-based	Affirmative	Negative
Mode of operation	Mandatory regulation	Standards	Fiscal (dis)incentive	Non-fiscal (dis)incentive	Encouraging behaviour	Warnings, moral suasion
Example(s)	Emission limits for coal-fired power stations	Green emission sticker for city centres	Emission trading systems, effluent fees, taxes	Award a certificate to the least waste producing company of the year in its branch	“Ride your bike to work week”	Public campaign on deforestation
HARD <<<		>>> SOFT				

Source: adapted from Table 1 in UK Science and Technology Committee (2011) and Figure 11.1 in Vedung (1998)

Each policy instrument is usually chosen, designed and implemented with a specific problem in mind (in this case an environmental concern), at a particular time and in the context of the current political-ideological situation of the government (Borrás & Edquist, 2013, p. 3). Nevertheless, the resulting uniqueness of each policy instrument does not impede a classification according to the general logic behind the political action (Borrás & Edquist, 2013, p. 3).

There are many different ways to classify policy instruments (e.g. Linder & Peters, 1998; UK Science and Technology Committee, 2011, p. 10). However, for environmental policies in particular, three main types of instruments can be distinguished: *regulatory*, *economic* and *information* (Mickwitz, 2003, p. 419; Vedung, 1998, p. 250). These instruments can be categorized in accordance with various criteria. One way is to arrange them based on the degree of the authoritative force involved (Mickwitz, 2003, p. 419). Table 1 shows the three types

of policy instruments, organised by their degree of restrictiveness from left (hard) to right (soft).

Regulatory instruments

Regulations (also known as command-and-control measures) can be defined as “measures undertaken by the governmental units to influence people by means of verbally formulated rules and directives which mandate receivers to act in accordance with what is ordered in these rules and directives” (Vedung, 1998, p. 51). In environmental policies, regulatory (restrictive) instruments can be subdivided into mandatory regulations and standards (Michalek et al., 2015, p. 10). Environmental regulations have been commonly used to intervene and eliminate environmentally damaging behaviour in most industrialised countries (Mickwitz, 2003, p. 419). Prominent examples of mandatory environmental regulation include legislative bans on activities that are harmful for the environment, such as an upper limit on emission levels for power stations burning coal in the US Clean Power Plan (CPP). In the case of other command-and-control measures, standards are introduced, e.g., prohibiting cars lacking a green emission sticker to enter the city centre.

Economic instruments

Economic instruments are tools based on the logic of the market and thus aim to create or alter benefits and/or costs for the agents involved (Mickwitz, 2003, p. 419). This economic category “encompasses tools that guide individual behaviour in a direction deemed beneficial by the regulator (e.g. environmental protection and/or sustainability) but *does not formally restrict* an individual’s freedom of choice” (Michalek et al., 2015, p. 10). Market-based instruments function by either handing out or taking away material resources. This can be done in-kind (non-fiscal) or in cash (fiscal) (Vedung, 1998, p. 51). The most prominent examples of market-based environmental policy instruments are the European Emissions Trading System (EU ETS) or the deposit system for cans and bottles that exists in some European countries such as Germany. A softer, non-fiscal tool that uses non-financial incentives is, for example, awarding certificates to best-practice companies that create the least amount of annual waste in comparison to their competition.

Information instruments

Information as a policy instrument “covers attempts at influencing people through the transfer of knowledge, the communication of reasoned argument, and persuasion. No more than the plain transfer of knowledge or persuasive reasoning is offered to influence people to do what the government deems desirable” (Vedung, 1998, p. 51). This can be done through moral suasion, exhortation or public communication. Information can be provided in an affirmative way, for example, by designating one week a year when everyone is encouraged to travel by bike instead of by car. Through this initiative, people learn from experience that more sustainable and less stressful alternatives to driving or the public transportation system exist. On the other hand, information campaigns can also foster more ecologically conscious behaviour by presenting negative information, for instance, warning people in a

public campaign about the impact that deforestation has on the planet to promote the reduction of paper waste.

In the last years, the use of soft instruments has been steadily increasing in many countries (particularly visible in Europe and the US). This development towards more integrative governance is at the core of a fundamental transformation that sees the role of government changing from that of provider and regulator to coordinator and facilitator (Borrás & Edquist, 2013, p. 4). In the course of this transformation, soft instruments have become an important part of the policy mix (Borrás & Edquist, 2013, p. 4). Green nudging falls into this category.

One characteristic common to the three conventional policy instrument types presented above is that they make the primary assumption that every individual reflects and considers all options available and finally makes a utility-maximising decision in line with stable preferences (Michalek et al., 2015, p. 11). This is partly because nearly all economic policy analyses use the traditional, neoclassical economic paradigm, assuming rational actors. In terms of the dual process theory of cognition described above and in Kahneman (2012), the traditional instruments rely on the reflective and self-aware decision making of System 2 to work as intended. Since the introduction of bounded rationality, it is clear why, so far, some of these policy tools have turned out to be virtually ineffective in solving environmental issues. This has as much to do with the nature of environmental issues as with the fact that decisions taken by individuals are determined by a multitude of factors in addition to prices, product information or bans on environmentally harmful goods and practices (Michalek et al., 2015, p. 2). As described above, research in behavioural economics has shown that heuristics and their related cognitive biases have a large influence on the decision-making process in terms of environmental issues – and that this occurs without any reflection on the part of the individual (Reisch & Hagen, 2011, p. 222).

3.2.2 Nudging in an environmental context

In Section 2, we introduced dual process theory, which claims a distinction between two different but interacting cognitive processes – one automatic, intuitive and driven by emotions (System 1), the other analytical, conscious and deliberately controlled (System 2). As per our technical definition of nudging in Figure 1, the mechanisms of a nudge function within the unconscious mode of System 1. Conversely, the traditional policy instruments used to foster environmentally friendly behaviour assume that humans rely principally on System 2 processes. This unique attribute of green nudging simultaneously sets it apart from and makes it a potentially key addition to the typical policy toolkit (Schubert, 2016, p. 3).

Here we introduce green nudging as a soft, non-regulatory extension of the environmental policy toolkit. In contrast to the commonly used instruments, presented above, nudging approaches human decision making in a completely different way. It relies on a more nuanced picture of human behaviour found in the behavioural sciences such as cognitive and social psychology and sociology. Research in these fields has identified a series of interventions that can be made to influence behaviour by changing decision context (Mont, Lehner,

Heiskanen, Sverige & Naturvårdsverket, 2014, p. 22). Nudging extends the ‘soft end’ of the toolkit, introducing an additional non-regulatory, potentially non-intrusive way of pursuing environmental ends. While nudge-based interventions belong under the category of non-regulatory instruments, they are not synonymous with all non-regulatory policy instruments. For instance, openly persuasive interventions and the straightforward provision of information do not count as nudges since the underlying idea of nudging is to prompt choices primarily without necessarily triggering a conscious consideration of the options available (UK Science and Technology Committee, 2011, p. 12).

Therefore, there are two fundamental differences that distinguish nudges from conventional instruments: (1) they do not assume the prevailing economic paradigm of rational choice and (2) they influence behaviour via System 1 processes – as per our technical definition. Neoclassical economic thinking, which is at the core of all conventional economic considerations, cannot be as easily applied to nudging. Meanwhile regulatory, economic and information policies are designed with the assumption of completely rational human beings who make all their decisions within a System 2 thinking mode. Nudging explicitly addresses automatic and intuitive thinking, which steers many of the decision-making processes in daily life. Thus, nudging goes beyond other soft policy instruments – such as information provision, which require deliberate thinking – into the sphere of automatic and intuitive decision making (Michalek et al., 2015, p. 11).

Unlike other conventional policy instruments, nudging influences individuals by either blocking or triggering the cognitive biases we all have and thus would not be effective on a totally rational actor (Schubert, 2016, p. 3). They can only function in a bounded rationality world, where individuals have limited cognitive capacity, attention and willpower and where preferences are ‘constructed’ instead of ‘given’ (Schubert, 2016, p. 3). However, this does not mean that System 2 processes are not affected by nudges. In Section 2, we point out how the two systems constantly interact with each other. For the analysis of nudging as a policy instrument this means that nudging first triggers an intuitive process, blocking or exploiting a cognitive bias. Subsequently, this process can (but must not always) affect a conscious, analytical assessment of the decision at hand (Hansen & Jespersen, 2013, p. 15).

Especially for environmental ends, nudging can be a very effective tool (Dolan, Hallsworth, Halpern, King & Vlaev, 2010). This is in part due to the nature of environmental problems. As discussed in Section 3.1, environmental problems are usually quite complex, span over long periods of time and their consequences and causes are often unequally distributed. Moreover, free-riding and the deprivation of public natural goods are often key challenges. The acknowledgement and integration of behavioural insights into the process of environmental policymaking marks a significant step towards a more effective regulation of environmentally harmful behaviour (Momsen & Stoerk, 2014, p. 376).

However, one aspect that must be considered when comparing and weighing the different policy instruments against each other is that several tools can also be used in concert. In

some cases, a good mix of different policy tools can help to address one issue more effectively than any one used in isolation (Schubert, 2016, p. 3). The UK Science and Technology Committee (2011) states as one of its central findings that “non-regulatory measures used in isolation, including nudges, are less likely to be effective. Effective policies often use a range of interventions” (p. 5). Although this might be a useful approach in practice, the analysis of the different policy instruments on a stand-alone basis is necessary to find out about the different characteristics and outcomes that each type of policy intervention has on its own. Only thereafter can research focus on potential mergers of several instruments to achieve a single goal. While this is not the object of the current study, it should be kept in mind when evaluating policy instruments in general.

Having covered the differences between nudging and conventional policy instruments, we turn now to a more detailed discussion of the various forms of green nudging. Not all green nudges function in the same way. The following section will shed light on the different types of green nudges, detailing their different characteristics as well as what implications these might have for an evaluation.

3.3 Towards a typology of green nudges

There is a growing body of literature on existing or proposed green nudges. First, we overview how past authors have categorised nudges before presenting a typology that complements our technical definition.

3.3.1 Categorising green nudges

Even a cursory review of the research on nudging will reveal that there is a multitude of ways that one could classify nudges. Most systems are grounded in a fundamental dichotomy between nudges that target solely the automatic processes of System 1 and those that help engage the reflective nature of System 2. For one, Hansen and Jespersen (2013) refer to these as Type 1 and Type 2 respectively, arguing that only nudges aimed at influencing System 2 processes can be called a manipulation of choice as System 1 only controls habitual behaviour. However, others have chosen to use a different nomenclature, such as ‘mindless’ versus ‘mindful’ (e.g. House & Lyons, 2013) or ‘heuristic-triggering’ versus ‘heuristic-blocking’ (e.g. Barton & Grüne-Yanoff, 2015). As we explain in Section 2, nudges function primarily via biases or systematic errors that arise *due* to cognitive heuristics – in other words, most of the time nudges do not trigger or block heuristics but rather biases. Hence, we find the ‘de-bias/re-bias’ dichotomy described in Amir and Lobel (2008) to be a more exact terminology. In the majority of cases, System 1 nudges influence behaviour by ‘re-biasing’ an individual, i.e., exploiting biases such as the propensity to stick with the status quo. Conversely, System 2 nudges generally ‘de-bias’ the individual, thereby facilitating active thinking.

Nudges can also be categorised using other metrics such as level of transparency or intended aim or target group. As for the former, Hansen and Jespersen (2013) distinguish between ‘transparent’ and ‘non-transparent’ nudges, stating that the difference resides in the degree

to which those being nudged can “reconstruct either the intention or means by which behavioural change is pursued” (p. 18). The authors combine this measure with their Type 1/Type 2 dichotomy to arrive upon a two-axis matrix for classifying nudges (see figure in Hansen and Jespersen, 2013, p. 20). House and Lyons (2013) further propose a method of categorisation based on whether the nudge promotes an individual’s own ends (e.g. quitting smoking, eating healthier), which they term ‘self-control boosting’, or whether it ‘activates behavioural standards,’ to which the individual is more likely indifferent (e.g. organ donation) (p. 7). This is somewhat alike the aforementioned ‘pro-self’ and ‘pro-social’ nudge categories (Hausman & Welch, 2010).

Nudges have also been grouped depending on how they alter subjective perceptions of utility. Within this framework, Fischer and Lotz (2014) describe four nudge types: (1) nudges that cause a reduction in perceived present (psychological or social) utility in favour of future utility thereby blocking the present bias (‘discomfort nudges’), (2) nudges that increase or decrease expected utility by influencing the subjective evaluation of probability (‘probability nudges’), (3) nudges that exploit indifference or holes in an individual’s utility function (‘indifference nudges’) and (4) nudges that influence behaviour not guided by utility evaluations, i.e., habits (‘automatism nudges’) (pp. 10-11). The authors admit that many nudges fall into two or more of the above categories as the underlying mechanism is commonly a mix of multiple different biases.

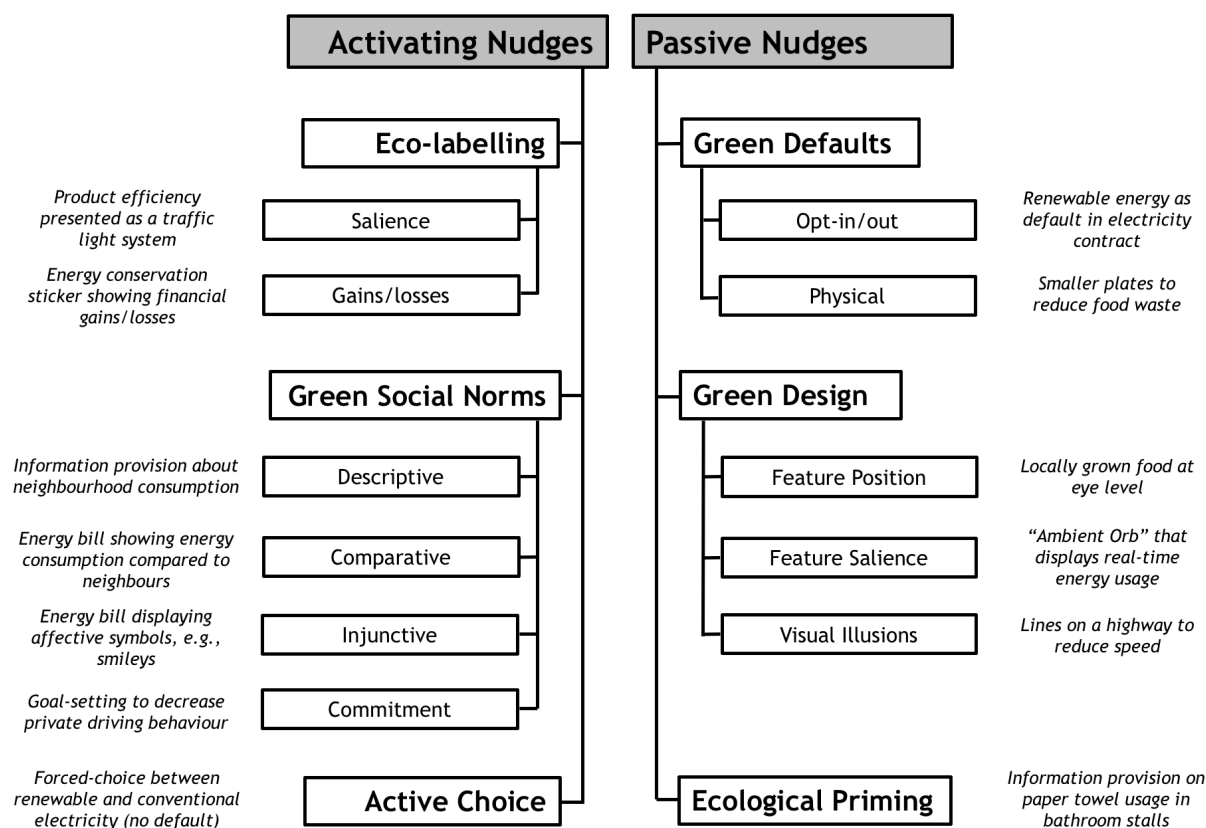
Finally, nudges can also be classified based on who is doing the nudging. *Public nudges* are mandated by a state or local government and directed at citizens while *private nudges* are employed in the private sector either by enterprising employers attempting to alter their employees’ behaviour while in the workplace or by companies maximising profit.

3.3.2 Proposed typology

Figure 2 depicts a proposed typology of green nudges developed for the purposes of this paper. The typology represents an amalgamation of many of the previously mentioned approaches to categorising nudges but leans heavily on the Type 1/Type 2 model used in Hansen and Jespersen (2013).

The typology uses a three-tiered hierarchical taxonomy system of increasing specificity. At the first level, *passive nudges* influence automatic, habitual behaviour and do not require any additional mental effort by the individual to work as intended. Conversely, *activating nudges* also make use of System 1 processes (e.g. social norms and framing) but steer the individual towards weighing and evaluating alternatives before making a more deliberate decision. Next, nudge types are grouped based on the underlying psychological mechanism. Finally, on the third tier, specific nudges are described using applied real-life examples.

Figure 2: Typology of green nudges



Green Defaults

Opt-in/out defaults refer to nudges that exploit the status quo bias by setting the desired ecological behaviour or environmentally friendly choice as the default. Above, we describe the example of preselecting the renewable energy option on home electricity contracts, but this nudge has also been used in consumer trials for home smart grids (Momsen & Stoerk, 2014; Ölander & Thøgersen, 2014) and attaching carbon offsets to travel purchases (Araña & León, 2013). Physical defaults involve changes to physical attributes of the choice architecture, which then set a new default by imposing boundaries or limits on available options without fully erasing them. A common example here is the reduction of plate size to combat food waste in cafeterias (for empirical examples refer to Kallbekken & Sælen, 2013; Wansink, 2004). Students can in theory return for seconds but tend to consume and waste significantly less food.

Green Design

Feature position refers to alterations in the way a choice architecture is arranged such that the ecological options are more noticeable (e.g. placing locally grown groceries at eye level in the aisle). Similarly, *feature saliency* aims to make ecological choices more noticeable but does so instead by emphasising specific aspects of the choice architecture; e.g., the

‘Ambient Orb’ product provides buyers with a real-time feedback on their energy use, turning red at high levels of consumption (Michalek et al., 2015, p. 5). *Visual illusions* make use of common visual perception tricks to induce, e.g., slower driving.

Ecological Priming

In *ecological priming*, individuals are primed with an informational message or visual stimuli relating to environmental protection or common environmental themes. The primed stimulus remains in associative memory and influences behaviour in a predictable manner over a certain period of time. For instance, in a laboratory setting individuals primed with a pro-environmental message were significantly more willing to buy package-free produce (Tate, Stewart & Daly, 2014, p. 230). Another commonly cited example includes environmental messages posted to restroom stall doors or above urinals, which aim to decrease paper towel usage after hand washing. Priming functions at a subconscious level but is not to be confused with subliminal messaging, which uses hidden stimuli that are not visible or noticeable to alter an individual’s choices. A priming cue may affect sub-conscious processes but is a conspicuous element of the choice environment (Dolan et al., 2010).

Eco-labelling

Eco-labelling refers to point of purchase messages such as product efficiency information or organic markers. A prime example of *salience eco-labelling* is the European Union (EU) energy efficiency labelling scheme for common household electrical appliances first introduced in 1995 (European Commission, 2012). The label is more than simply information provision because it affects the consumer at the point of purchase and thus counts as an element of the choice architecture. Moreover, the design of the label – its traffic light colour scheme and seven class hierarchy – has an impact on choice due to framing effects as well as the anchoring heuristic (Ölander & Thøgersen, 2014). Eco-labelling can also be framed in terms of gains and losses, e.g., the amount of money saved by switching to a more efficient appliance. *Gain/loss eco-labelling* encourages more energy efficiency and sustainable consumption due to loss aversion and naturally framing effects. If presented in terms of lifetime costs or savings, energy efficiency labelling can help overcome present bias and hyperbolic discounting (Kallbekken, Sælen, & Hermansen, 2013). Eco-labelling also functions via implicit social norm framing. That is, the ubiquity of appliance efficiency tags or fuel rating stickers on cars speaks to the green social norms that exist in a society. Thus, eco-labelling may function not only through salience and framing effects but also by appealing to a buyer’s self-image and desire to belong to an ecologically-conscious society (Schubert, 2016).

Green Social Norms

There are three kinds of social norm nudges: descriptive, comparative and injunctive (Cialdini et al., 1990). *Descriptive social norms* provide information on what is generally done, i.e., information describing the behavioural norm. For instance, Schultz et al. (2007) provided 290 households with feedback on the average monthly electricity consumption in their immediate neighbourhood. This information was displayed directly alongside the usage

of each individual household. Over the span of one bill cycle (a month), the *comparative social norm* induced an average reduction in electricity consumption of 1.22 kWh daily for those households above the neighbourhood average. Intriguingly, those below the neighbourhood average tended to increase their electricity consumption. This *boomerang effect* in energy usage was diminished when the researchers added an emoticon to the electricity bill. If usage stayed below average residents received a smiley face; residents above average received an angry face. The emoticons evoke an *injunctive norm*, i.e., that what is morally correct or right in society. Similar studies have shown that employing green social norms can decrease littering (Cialdini, 1990), prompt hotel guests to reuse their towels (Cialdini, 2005) and encourage national park visitors to refrain from taking home souvenirs they find on the trail (Cialdini et al., 2006). All social norm nudges employ *herding effects* or the natural propensity among humans to cave to societal pressure: “If everybody is doing it then it must be the sensible thing to do” (Cialdini et al., 1990, p. 1015). An injunctive norm further appeals to a perceived risk of *societal sanctioning* as well as *reciprocal altruism* (Ostrom, 2003).

Interestingly, the effectiveness of green social norms is mediated by a variety of different variables. For example, Costa and Kahn (2013) provide evidence to suggest that a conservative political ideology will decrease the influence of energy conservation social norms. In a laboratory experiment Graffeo et al. (2015) found that social distance, i.e., whether or not the norm described nearby neighbours or members of a broader population, has a negative effect on the nudge’s success. In other words, descriptive social norms seem to be more effective if they involve peers in the same general area or with shared circumstances, a finding that has been replicated by Goldstein, Cialdini and Griskevicius (2008).

Commitment nudges, such as setting an energy conservation goal (e.g. McCalley & Midden, 2002), function in part due to the social pressure involved with making promises, the desire to remain behaviourally consistent and the social power of reciprocity (Cialdini, 2007, Chapter 3). Failing to fulfil a commitment also induces a state of cognitive dissonance, which we actively avoid by aligning actual behaviour with behavioural intent and attitude. Experimental research suggests that even self-imposed commitments made in private can positively impact behaviour (Dolan et al., 2010, p. 26). For instance, augmenting a goal with an *implementation intention* – i.e., designating a time and place to fulfil the goal – can increase use of public transport (Bamberg, 2002).

Active Choice

Active choice or prompted choice nudges purposefully frame decisions without including a default. This activating nudge forces the nudgee to deliberately pick from various options. Active choice is pushed by some academics as a less controversial and thus more ethically palatable alternative to strictly opt-out framing nudges such as green defaults (see for instance Gigerenzer, 2015).

It is important to note that not all nudge examples fit perfectly into the typology outlined above and some could arguably fall under two or more of the six proposed categories. For instance, while not commonly used for environmental aims, *simplification nudges* are design nudges that frame complicated decisions in simple terms. Simplification is thus an activating (or de-biasing) nudge because it helps individuals avoid cognitive fatigue and the likelihood of falling prey to a System 1 bias (European Commission, 2016, p. 16). Similarly, large-scale changes to the choice architecture – e.g., a higher collection frequency for kerbside recycling than trash – also fall under the heading of green design but fall just outside our typology.

3.4 Should we nudge for the environment?

At this point, we have covered the theory and background of nudging, embedding this within the context of environmental policy. Furthermore, we presented a typology of green nudges, describing each nudge with examples of its applied use. At face value, it may seem that green nudges are well suited to address environmental problems and in some cases may be more promising than conventional policy instruments. Nevertheless, policymakers require an evaluative framework in order to adequately assess the potential of nudging to meet environmental goals. This framework must consider practical aspects such as the efficacy and effectiveness of the instrument, but also the many ethical considerations that pertain to public policy-induced nudging. Indeed, as a unique addition to the typical environmental policy toolkit, green nudging cannot be readily evaluated using traditional strategies. Evaluating green nudges calls for new lines of thinking, which reflect the numerous disparities between nudge-based policies and conventional policy instruments.

Before we turn to the evaluative framework, it is important to say a few words on our normative approach going forward. In this paper, we ask the question: Should we nudge for the environment? We have decided to focus our operations on green nudging as a *means*, a mechanism or tool to induce sustainable behaviour and lifestyles. This normative stance assumes the ends of green nudging to be justified; that is, we assume that achieving environmental policy goals is broadly accepted as a good thing for society. We contend that this precondition is at least partially justified because countries all over the world have pursued environmental aims for decades, with increasing intensity – a trend that is likely to continue into the foreseeable future with more and more emphasis on the international level (e.g. UN institutions such as the recently adopted Paris Agreement). Moreover, political momentum is usually mirrored by public acceptance particularly in environmentally progressive countries, like Germany, where nearly one in five individuals sees environmental and climate protection as the number one problem facing society and policymakers (BMUB, 2017, p. 15). The same survey research found that 66% of the German population demand that the government do *more* to protect the environment (BMUB, 2017, p. 22). Finally, if we begin to doubt the assumption that environmental protection is a desirable end, this would mean that we must question the worth of all environmental policies, not just green nudging. Our normative stance operationalises nudging as a means and thus enables us to omit any further

discussion of the political philosophy of libertarian paternalism, insofar as it pertains to questions about desirable ends. Nevertheless, we are aware that the aforementioned assumptions are debatable and accordingly some of these issues resurface in our discussion of ethics in Section 4.2.3.

4 Research design

In the following, we outline this paper's research approach. First, the conceptualisation and methodology are presented, which includes necessary background information about policy evaluation in general as well as an introduction to our evaluation criteria. Second, our evaluative framework is explained in more depth, drawing on findings from a comprehensive literature review to illustrate the importance of each criterion and sub-criterion. Third, we justify the choice of our criteria, noting possible limitations. With the development of this evaluative framework it is our hope to contribute to the early stages of the implementation of a policy that has the potential to augment and change environmental policymaking.

4.1 Conceptualisation and methodological approach

Policy evaluation represents an important stage in the traditional policy cycle and facilitates the analytical structuring and better understanding of political processes (Jann & Erich, 2009, p. 97). In the specific field of environmental policies and environmental policy instruments, several distinct evaluation approaches exist. Depending on the evaluator's intentions, different approaches serve different purposes. A side effect evaluation for instance, is a "qualitative categorization of effects" and allows evaluators to look at "anticipated and unanticipated effects" of a policy as well as whether these effects "occur inside or outside the target area" (Hilden et al., 2002, p. 15). It is thus a useful tool for analysing an established environmental policy instrument. For our purposes, side effects play a role because in the contested field of nudging, anticipated and unanticipated effects can occur. Any investigation of unanticipated effects, however, will be limited due to the fact that our evaluation only represents a preliminary 'glimpse' of two types of nudges. Furthermore, only a limited evaluation of unanticipated effects is possible since the development of green nudging in the public policy realm is still in its early stages.

Another relevant type of evaluation, which serves as a valuable base for our research, is the evaluation of 'recently introduced policy instruments' (Mickwitz, 2003, p. 421). Nudging as a public policy instrument in the environmental realm has just recently been introduced and thus requires an analysis that differs from ex ante or ex post evaluations. An ex ante policy evaluation assumes that no empirical tests of available experiences exist whereas ex post policy evaluations require data regarding the outcome and unanticipated effects (Mickwitz, 2003, p. 421). Neither holds true for a recently introduced policy instrument such as nudging. In this sense, our approach differs from an integrated policy instrument evaluation. We do not conduct our own empirical tests and produce data but rely on primary literature to provide us with the relevant information. Due to this, we do not intend to focus on efficiency or cost-effectiveness/cost-benefit analysis. Even though we agree with the statement that

economic criteria are important and should be part of an in-depth, integrative policy instrument evaluation (Hilden et al., 2002, p. 16), it does not seem suitable to include this in our evaluative framework. Data limitations make it difficult to acquire the knowledge needed to give adequate recommendations this early in policy development. Therefore, we decide to focus instead on three criteria linked to the democratic realm of environmental policy and which we deem to be specifically important for nudging. Finally, in this paper we take a proactive evaluation approach, meaning that our research is conducted to “serve decision makers” planning by “delineating, obtaining and providing information” (Gave-May & Pal, 1999, pp. 265-266).

In terms of methodology, our evaluation consists of a literature review embedded in a conceptual frame deduced from different environmental policy evaluations. The preliminary evaluation we perform focuses solely on two types of nudges (green social norms and green defaults) rather than on green nudging more broadly. As researchers, we recognise the fact that politics and values go hand in hand with the evaluation of public policy. As Gave-May and Pal (1999) stated “data have to be interpreted and filtered through values of some sort that help us determine what is significant and what is not” (p. 260). While we understand this limitation, a full treatment of these post-structuralism concerns is beyond the scope of this research project.

4.2 Evaluative framework

In this section, we start by providing an in-depth overview of our evaluative framework and its three main criteria – *efficacy*, *effectiveness* and *ethics*. All three criteria and associated sub-criteria are presented in Table 2 and are subsequently defined and discussed.

4.2.1 Efficacy

Before implementing a new policy instrument, social scientists as well as policymakers often underestimate the importance of efficacy. In fact, while every stakeholder involved in the implementation of new social policy should be aware of the necessity of sound research and testing, this is not always the case. In the past, in the field of public policy, decent efficacy trials have often been neglected, with policymakers introducing a new policy without knowing if the policy does in fact work. This failure to incorporate efficacy has several explanations, such as the complexity of society and the problem and challenge of generalising on a large scale. Another possible reason has to do with practical feasibility; it is not possible to conduct experimental trials for each new policy suggestion. However, due to the complexity of environmental problems and the empirical foundations of nudge theory, we contend that it is vital and possible to test green nudging as a new type of policy instrument in regard to its functionality and efficacy.

The efficacy of a public policy instrument is confirmed by scrutinising interventions under optimal conditions in a lab setting and evaluating whether the positive effects of a programme outweigh its potential harms, or as Flay et al. (2005) state: “Efficacy refers to the beneficial effects of a programme or policy under optimal conditions of delivery” (p. 3). The

standard scenario or efficacy test of a social policy uses randomised experiments or trials. Several alternative techniques, such as case studies or analyses of statistical evidence are used in the field of social science as well. However, as mentioned by John et al. (2013) “experiments – especially *Randomized Controlled Trials* [RCT, emphasis added] – offer a much more valid and robust standard of evidence than any other method policymakers have available” (p. 28). Particularly in the fields of behavioural economics and experimental psychology, experiments are the standard method of inquiry. Experiments test the effects of a new policy instrument or policy interventions on a small population subset, or *sample* (Vo-year, 2015, p. 10). While running experiments, it is of great importance to keep in mind the drawbacks of generalisation and the importance of validity. In their book *Nudge, Nudge, Think, Think*, John et al. (2011) differentiate between three types of validity: *general validity*, *external validity* and *internal validity* (p. 29). Table 3 demonstrates the meaning and the differences between the three, which should be taken into consideration by any policy-maker or evaluator.

Table 2: Evaluative framework for green nudging

Criteria	Sub-Criteria
Efficacy	<ul style="list-style-type: none"> • Theoretical basis • Reliable methods (<i>internal validity</i>) • Replication (<i>external validity/general validity</i>)
Effectiveness	<ul style="list-style-type: none"> • Generalisability • Practicability • Longevity • Outcome/Impact • Side Effects
Ethics	<ul style="list-style-type: none"> • Autonomy • Manipulation • Transparency • Proportionality

The importance of efficacy in the context of policy implementation is out of question. Criteria by which to measure efficacy in the context of green nudging, however, have not yet been fully developed in the literature. The current research defines and measures efficacy, pulling from existing literature in a diverse array of fields. Due to the complex nature of environmental problems and hence environmental policy, conditions of pure efficacy trials might turn out to be more difficult than in other fields. This fact has to be kept in mind during future efficacy trials (Mickwitz, 2003, p. 416). In this paper, we divide the efficacy of green nudging into three sub-criteria based on the three dimensions of validity (see evaluative framework overview in Table 2): *theoretical basis*, *reliable methods* or *internal validity* and *replication* or *external/general validity*.

As a theoretical basis, there are four stages of research that Flay (1986) proposes for a programme or policy to be well planned. Although he focuses on health and education programmes specifically, we argue that the four stages are useful to plan a green nudge efficacy trial as well. Stage (1) is as simple as conducting basic research on the policy field of interest and similar existing programmes and is followed by stage (2), the development of a hypothesis. Once sound theoretical and empirical ground is laid, Flay (1986) advocates for the provision of studies in the form of applied (3) pilot research and (4) prototypes (p. 459).

Table 3: Three types of validity for the efficacy of policy interventions

General Validity	The extent to which the findings of an experiment or line of research can be generalised to different contexts and circumstances other than those of the original study.
External Validity	The extent to which it is possible to make inferences from an experiment or line of research.
Internal Validity	The extent to which it is possible in an experiment or line of research to conclude a hypothesized relationship or finding, i.e., whether it is possible to explain empirical findings.

Source: John et al. (2011), p. 29.

An answer to the question of whether an intervention is efficacious or not can only be given with the help of different kinds of reliable methods. Depending on the circumstances, these methods consist of different types of experiments. Experiments can be conducted in a laboratory setting or under real-life conditions. Efficacy, however, only concerns experiments conducted in the laboratory or under highly controlled, laboratory-like circumstances. The natural and social sciences differentiate between two types of experiments: *true experiments* and *quasi experiments*. A true experiment requires the random selection of participants and the random assignment of these participants to either control or treatment groups. Furthermore, true experiments are characterised by the manipulation of independent variable(s) in order to measure the outcome and function and to show cause-and-effect relationships. While making a hypothesis based on a theoretical literature review, scientists manipulate the hypothesised independent variable to affect the outcome. This action is typically referred to as an intervention. Quasi experiments do not assign their participants randomly because it is either not possible or practical, such as is the case when investigating differences between demographic groups (Gribbons & Herman, 1997, pp. 1-5). Before running a scientific experiment or trial, one of the main conditions for scientifically sound data and information in efficacy processes is the obtainment and specification of the sample. Who is part of the randomised sample group and how well do they represent the affected population of a policy instrument? As stated above, generalisation and the problem of validity have to be taken into consideration while conducting efficacy trials. When thinking about

reliable methods and replication, policymakers and evaluators have to be aware that experiments or trials differ when it comes to their internal or external/general validity (John et al., 2011, p. 36).

Nudging, as a newly introduced environmental policy instrument, is an interesting case since trials might not be possible for every type of environmental nudge at the early stage of the policy. Sometimes it proves difficult to run fully randomised trials and therefore quasi experiments in the form of *design experiments* might be useful and appropriate for policymakers and evaluators. Design experiments come into play if there is still little known about an intervention. Such experiments serve as the foundation for future RCTs as they give decision makers an overview of what works. Anyone who conducts a design experiment has to deal with the question: *What is the point of my study?* (Cobb, Confrey, diSessa, Lehrer & Schauble, 2003, p. 11). John et al. (2011) define design experiments as a “qualitative experiment, which focuses on the design of an intervention as the thing to get right. The experimental aspect of the method manipulates an intervention and observes it over an extended time period, usually in one location, until acceptable results emerge” (p. 38).

RCTs are usually considered the “gold standard [...] that offer a genuine counterfactual of what would have happened without the intervention [...]” (John et al., 2011, p. 33). A sample group is randomly selected from a population of interest and categorised randomly into intervention and control groups. Randomisation is essential to ensure similarity in all respects and account for possible confounding variables – i.e., randomisation ensures that co-varying factors are spread between the control and treatment conditions. For an intervention to be considered efficacious and valid, the sample must show statistically unbiased estimates and no serious negative effects on the outcome can occur (John et al., 2011, p. 32). Each experimental result must be reported accurately and consistently, regardless of whether it is positive, negative or non-significant. After the initial experiment, the same manipulation or intervention can be repeated at a later point in time with a new sample. This procedure is called *replication*. Replication is an important efficacy tool, as it adds to the external/general validity and generalisability of an experimental finding. The person who conducts the replication need not be the same researcher. Additionally, the replication can either copy the original study by holding all conditions unchanged or attempt to show that the empirical findings are valid under multiple conditions (i.e. for cross-cultural studies). Here, however, the line between tests of efficacy and effectiveness becomes very fine and often overlaps, since replications can be conducted under real-life settings as well. Such *natural experimentation* or pilot testing investigates the effectiveness and not the efficacy of a policy intervention.

Efficacy lays the foundation for an effective public policy. Ideally it contributes to careful research and experimentation. Effectiveness in the real-world setting, however, requires a different set of evaluative criteria, which we discuss in the following section.

4.2.2 Effectiveness

One of the central foci in the emerging academic literature on environmental policy instrument evaluation and on behavioural policy tools is effectiveness (Glasgow, Lichtenstein & Marcus, 2003; Mickwitz, 2003; Schubert, 2016; Steg & Vlek, 2009). Mickwitz (2003) underlines the European Union's emphasis on effectiveness in environmental policy evaluation, highlighting the Environmental Action Programme for the European Union (1600/2002/EC), Article 10 (European Council, 2002). "Possible impacts" on the environment and "effectiveness of existing measures" are to be evaluated by policymakers (Mickwitz, 2003, p. 416). But what exactly does effectiveness mean in this context? What criteria can be used to measure the effectiveness of an environmental policy tool such as a green nudge?

What efficacy and effectiveness have in common is that they are both used to evaluate whether a policy intervention leads to the desired outcome. However, efficacy is usually tested first to identify whether the policy works in a lab setting. In a second step, the impact in the real world must be assessed. A programme cannot be considered effective if tests for efficacy fail to detect a positive effect in the first place, meaning that interventions have to be shown to do more good than harm in order to be suitable for an effectiveness evaluation (Flay, 1986, p. 454). Efficacy and effectiveness are therefore closely connected and account for a significant part of a policy evaluation. As explained by Nathan, Stuart and Dolan (2000), "Efficacy studies emphasize internal validity and replicability; effectiveness studies emphasize external validity and generalizability" (p. 965).

The question asked in this context is whether a policy produces the intended effects within a broader population (Glasgow et al., 2003, p. 1262). Flay et al. (2005) add to the definition of effectiveness by describing it as the "level of good over harm (or benefits over costs) that a programme achieves when received under typical real-world conditions of availability and acceptance" (p. 468). Steg and Vlek (2008) point out that there is a research gap in the evaluation of policy interventions other than informational strategies. Thus, this holds true for behavioural policy instruments such as green nudges. According to Steg and Vlek (2008), these instruments need to be evaluated for effectiveness as well, since they are potentially important tools to enhance pro-environmental behaviour.

In conclusion, effectiveness as a criterion in policy evaluation plays an important role and can be defined as coherence between intended results and actual outcomes. Yet what criteria can policymakers use in order to ensure that a green nudge is effective? An effectiveness trial can be called successful only if the policy has robustly gone through a range of settings with various participants, conditions and audiences, resulting in the intended outcomes. Glasgow et al. (2003) have listed a spectrum of characteristics that play an essential role in effectiveness trials. A policy instrument is effective if it reaches an extensive part of the population, if it can be applied in different social settings (for example both at work and in schools) by different trained staff and if it has long-term effects that are replicable. Moreover, negative side effects should be kept to a minimum and costs should be low (Glasgow et al., 2003, p. 1264). Similarly, Steg and Vlek (2008) suggest that the effectiveness of

policies with single or multiple intervention strategies should be researched through experimental research designs that include control groups and monitor changes in behaviour over a long period (p. 314).

According to Flay et al. (2005), a successful effectiveness trial is characterized by a list of criteria that add to a successful efficacy trial. In line with the above given criteria by Glasgow et al. (2003), one important principle is to facilitate the trial in a way that third parties are able to undertake the policy implementation. Furthermore, evaluation must take place under real-world conditions, involving the tested population groups in the intervention and its control. In addition, the practical relevance of the outcome needs to be demonstrated and explanations of how the outcome for a specific group can be generalised should be given (Flay et al., 2005, p. 19). With regard to methodological standards in effectiveness trials, Flay et al. (2005) maintain that they should be the same as in efficacy trials, with a preference for randomisation (for example RCTs). However, the authors acknowledge that maintaining the same standards is a challenge in real-world situations. Nevertheless, it is desirable to establish high quality effectiveness trials, which for instance can be achieved through a pilot stage in which the initial policy developer is not involved. Avoiding the use of resources that are not available in reality is a criterion that ensures practical importance of the effectiveness trial (Flay et al., 2005, pp. 12-14).

One tool to evaluate the effectiveness of nudging as an environmental policy instrument is the *field experiment*. As with the discussion of efficacy experiments above, it is important to distinguish between true experiments and quasi experiments. Field experiments often count as true experiments outside of laboratory settings. They provide policymakers with a useful tool to conduct sound research and tackle policy questions, since they differ from laboratory experimentation. As the name already suggests, field experiments consist of experimentation and fieldwork. The experimentation strategy constitutes random assignment of observed units to treatment or control groups. As with RCTs, randomisation ensures equal starting conditions and equal expectations. Field experiments are similar to natural experiments in that both occur outside the laboratory. However, natural experiments are considered quasi experiments conducted outside of a laboratory setting – thus randomisation and independent variable manipulation are not always possible. Gerber and Green (2011) describe natural experimentation as: “[representing] a deliberate departure from observational investigation, in which researchers attempt to draw causal inferences from naturally occurring variation as opposed to variation generated through random assignment” (p. 1). John et al. (2013) argue that external validity might be weak due to a limited scope of trials and the lack of experimental control (pp. 34-35; see also Pichert & Katsikopoulos, 2008, p. 66). However, this does not necessarily hold true if trials are replicated at different times, within different contexts and for different populations, facilitating the accumulation of knowledge.

Two additional aspects of a green nudge's effectiveness are that it should be generalisable and it needs to show long-term effects. Policy developers aim to see long-term effects attributed to green nudges, but these are tied to another condition: The achieved outcomes of a nudge need to correspond to the intended goals (Mickwitz, 2003, p. 426). As Flay et al. (2005) put it, the policies have to be "ready for dissemination so that they can be implemented effectively, that is, in a manner that achieves the expected effects" (p. 3). Thus, a nudge is only effective if it results in the anticipated outcome. The effectiveness trial helps to test implementation quality and consequently fosters the achievement of an anticipated and desired outcome (Flay et al., 2005, p. 3).

A problem in this context is the occurrence of unanticipated side effects. Already in the efficacy trial it is recommended to measure potential side effects and to observe negative outcomes unrelated to the intended ones (Flay et al., 2005, p. 6). The effectiveness trial is the essential next step because in the real world, there may be outcomes that the policy developers did not foresee. This holds especially true for green nudges and the environmental policy domain, since they aim for a positive outcome in a very complex system where effects often become apparent only after a long period of time. Glasgow et al. (2003) agree with, and underline, Flay's concern for the "measurement of potential harmful outcomes", which in their opinion has not sufficiently been addressed in the testing of effectiveness (p. 1264).

Besides the focus on the outcome of a green nudge, another effectiveness criterion that we briefly mention above is its practical relevance. This is closely connected to the nudge's impact. What Flay et al. (2005) claim for health policies is just as relevant for environmental policies: Interventions can only have a significant impact if they are relevant in practise (p. 11). Therefore, it is not enough to test a nudge in the theoretical framework of an efficacy trial, but to additionally assess it in different settings in the real world. Flay et al. (2005) further suggest reporting on the impact of the findings during the effectiveness trial. The practical value of the impact can amongst other things be measured in the form of standardised effect sizes, confidence intervals or odds ratios (Flay et al., 2005, p. 15). In their chapter *Testing*, John et al. (2011) point out the importance of the evaluation of impact to ensure effectiveness as well. They propose the use of conventional social sciences methods, for example case studies and statistical analysis or comparisons, to evaluate the impact of a nudge. According to the authors, RCTs are the most powerful method to investigate whether a nudge has the intended impact.

To sum up, we have collected a selection of criteria that assure a sound effectiveness trial. If we want a green nudge to be effective, the precondition is that it must successfully pass the efficacy trial. In the actual effectiveness trial, various other criteria need to be considered. As an agglomeration of the scholarly definitions above, effectiveness in our evaluative framework includes the following principles (Table 2): *generalisability* (can the audience, setting and implementation be applied beyond an initial experiment?), *practicability* (is effectiveness tested in a real-world setting?), *longevity* (long-term effects), *outcome/impact*

(does the nudge lead to the desired behavioural outcome and a positive impact on the environment?) and *side effects* (unanticipated and anticipated).

4.2.3 Ethics

In the case of nudging and specifically in the field of green nudging, ethics plays a significant role even though it falls by the wayside in standard evaluative approaches. A more integrative evaluative frame would serve future evaluators and policymakers in discussions on whether to employ green nudging. As discussed above, green nudging does not fit into the traditional policy toolkit, primarily because its mechanisms function within the unconscious mode of System 1. Purely focusing on efficacy and effectiveness is, therefore, not sufficient in answering the question: Should we nudge for the environment? In this section we provide a summary of the leading ethical criticisms in order to provide a necessary overview of the current state of debate. The ethical arguments discussed below, form the basis for our evaluative criteria, which is summarised in Table 4.

Table 4: Criteria for an ethical evaluation of green nudges

Autonomy	The capacity to be one's own person, to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external forces
Manipulation	The attempt to influence people subconsciously or unconsciously, in a way that undermines their capacity for conscious choice
Transparency	The intention behind a nudge, as well as the means by which behavioural change is pursued, is apparent to the agent being nudged
Proportionality	Must fulfil four elements: serve a legitimate aim; nudge needs to be suitable; must be the least restrictive means; nudge needs to be adequate

To start, nudging has been widely criticised on the basis that it infringes upon one's own *autonomy*. We understand autonomy to "[...] refer to the capacity to be one's own person, to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external forces" (Christman, 2015). Fischer and Lotz (2014) discuss the ethical legitimacy of behavioural policy measures in their report *Is Soft Paternalism Ethically Legitimate?*, identifying freedom of action and freedom of will (autonomy) as vital relevant ethical criteria in the discussion of nudging as a policy instrument. Smith, Goldstein and Johnson (2013) further argue that the focus should be placed on consumer welfare, more specifically to what degree it is maximised in relation to the level of loss of autonomy (p. 161).

Schubert (2016) conducts an ethical assessment of green nudges, offering some critical guidelines that may help practitioners come to an ethically informed assessment. The author argues that central to an ethical assessment of green nudges is the level to which it compromises people's autonomy, its impact on 'self-legislation' and the fairness of the nudge (p. 5). He proposes that autonomy considerations should enter the overall ethical assessment of green nudging as one kind of 'normative cost', to be weighed against benefits in terms of, for example, welfare. He concludes that transparency, coupled with the condition that the nudge generate durable behavioural change, is a key prerequisite for effective and ethical green nudging (Schubert, 2016, p. 29).

Although philosophers Hausman and Welch (2010) share Thaler and Sunstein's (2008) opinion that nudging (e.g. the establishment of default settings) does not restrict freedom of action or reduce individual utility, they do however argue that nudging remains harmful. Hausman and Welch (2010) hold that autonomy, defined as the deliberative process that brings about an individual's own preferences, is infringed. The authors conclude that it is problematic when decisions reflect the tactics employed by the choice architect because the degree of control over one's own mental processes is therein reduced, leading to preferences that in a certain sense are not one's own. Here the problem is not regarded as restricting a person's freedom to act, as Schnellenbach (2012) reasons, but rather as restricting freedom of will. Nudges can, however, promote as well as violate people's autonomy. This particular phenomenon will be discussed in more detail in the evaluation section of this paper.

One key component that emerges from contemporary philosophical autonomy theory is what Dworkin (1988) calls the "independent procedure requirement" (p. 18). The intricacies of this requirement cannot be discussed here in detail; suffice it to say that it holds that the process of arriving at a certain choice must be sufficiently independent from external influences. In contemporary society, however, our decision will always be subject to a significant degree of external input. Therefore, what seems crucial is that these influences be made explicit so that the subject can take a critical stance toward them. This concept also claims that transparent nudges do in fact generally respect autonomy and are necessary to ensure the safeguard of autonomy and democratic decision making (Alemanno, Sibony & Sunstein, 2015).

In the literature, the term *manipulation* is used both as a value-neutral, technical concept to simply mean 'intentional change' and as a value-laden, normative term. While the former aims to understand manipulation as a specific type of influence (putting aside the question whether it is good or bad), the latter assumes that there is something inherently wrong with manipulation. In this view, manipulation is primarily wrong because of the implied violation of autonomy.¹⁰ Nudges (those being nudged) are not in full control of their behaviour. They

¹⁰ Conly (2013) explains further: "Libertarian Paternalism is manipulative. That is, it does not suggest that we engage in free and open discussion in order to rationally persuade

are not doing the choosing themselves, there is no real self-determination and, hence, their autonomy is undermined. Wilkinson (2013) includes the potentially wrong-making qualities of manipulation:

“Manipulation is intentionally and successfully influencing someone using methods that pervert choice [...] Nudging clearly could satisfy the condition of having an intentional actor who successfully influences someone’s decisions. In principle, nudging could also pervert someone’s decision-making process and thereby infringe upon his or her autonomy. But if and when nudging does not use a perverting method, it is not manipulation. If and when nudging is not a form of intentional influence, it is not manipulation. If and when nudging does not succeed in altering behavior, it is not manipulation” (p. 347).

Even if we consider manipulation wrong because it impairs an individual’s autonomy, this does not necessarily mean that all nudges are manipulative. After all, a nudge can fail to meet one of the necessary conditions. In turn, this is not to say that a non-manipulative nudge is perfectly justified, as it can also be wrong for other reasons. Nevertheless, Nys and Engelen (2016) argue that there can be instances of justified manipulation (pp. 4-8).

Sunstein (2015), on the other hand, argues that an action does not count as manipulative merely because it is an effort to alter people’s behaviour (p. 443). Although he does describe that when one is manipulated, one is being treated as a kind of “puppet on a string”, and further states that no one wants to be someone’s puppet, and it is especially bad to be a puppet of government (Sunstein, 2015, p. 443). The author concludes that even with full transparency, at least some degree of manipulation may be involved whenever a choice architect is targeting emotions or seeking a formulation that will be effective because of how it interacts with people’s intuitive or automatic thinking (System 1) (Sunstein, 2015, p. 447).

However, it is frequently claimed that nudges generally and regularly count as manipulative. Conly (2013), for instance, believes that when we nudge we no longer regard people as generally capable of making good choices, and we out manoeuvre them by appealing to their irrationality, just in more fruitful ways (p. 30). She further argues that “we concede that people do not generally make good decisions when left to their own devices, and this runs against the basic premise of liberalism, which is that we are basically rational, prudent creatures who may thus, and should thus, direct themselves autonomously” (Conly, 2013, p. 30).

you to change your ways. [...] The point of the nudge is to push you in ways that bypass your reasoning. That is, they use your cognitive biases, like your tendency to go with the default option, to bring about good effects. There is a sense in which they fail to respect people’s decision-making ability [...]” (p. 30).

Ethical criticisms often focus on the claim that it is morally relevant *who* is doing the nudging. White (2013) gives two reasons why it is problematic for governments to nudge us (p. 51). First, there is the information and epistemic issue of whether government agencies and bureaucrats can accurately know our reflective judgements. Why are the manipulators in a better position than the manipulated to know what people really want? Why should the manipulators be trusted with this kind of judgement and power? Secondly, while we expect such manipulation from private companies, we do not expect it from governments and this can make us extremely vulnerable to it (White, 2013, p. 51).

For the purpose of our evaluation, we will be using Sunstein's (2015) definition, where a nudge is "manipulative if it attempts to influence people subconsciously or unconsciously, in a way that undermines their capacity for conscious choice" (p. 443). We will, however, be evaluating the level of manipulation through a framework established by Hansen and Jespersen (2013). The framework was established to serve as a tool for the responsible use of the nudge approach to behavioural change, and evaluates whether a nudge is manipulative or not, by examining how transparent it is, as well as whether the nudge pertains to a choice (reflective thinking) or not (Hansen & Jespersen, 2013, pp. 23-27). *Transparency* is particularly critical when evaluating nudging in the public sector, since it lacks the usual safeguards that accompany law making (White, 2013). Sunstein (2015) himself has addressed the need for transparency, stating, "an action counts as manipulative if it lacks transparency" (p. 443).

Hansen and Jespersen (2013) maintain that standard policy tools are seldom transparent, arguing that a transparent nudge is when the citizen being nudged, reasonably understands the intention behind it. A non-transparent nudge works in a way that the citizen being nudged cannot perceive the means by which the behavioural change is pursued (Hansen & Jespersen, 2013, p. 17). Hansen and Jespersen's (2013) perspective is contested, notably by Lepenies and Malecka (2016), on the basis that citizens – although they may understand they are being nudged – are likely to not fully comprehend the issue in its entirety: A nudge is the outcome of a process designed to influence citizens behaviour in order to achieve a certain policy outcome. Lepenies and Malecka's (2016) perspective thus renders all nudges as non-transparent, and as such, for the purpose of our evaluation, we will use Hansen and Jespersen's (2013) definition: A "[...] *transparent nudge* is defined as a nudge provided in such a way that the intention behind it, as well as the means by which behavioural change is pursued, could reasonably be expected to be transparent to the agent being nudged as a result of the intervention" (p. 17).

Furthermore, interventions need to be proportionate to the gravity of the behaviour and corresponding impacts they are trying to change. The legality of nudging has received some attention in the literature (e.g. Alemanno et al., 2015; Lepenies & Malecka, 2015), but the question remains as to whether governments may be obliged to nudge for environmental purposes. Although the evaluation of proportionality may not traditionally fall under the umbrella of an ethical evaluation, we have decided to include it here since we deem green

nudges to serve justified ends. Therefore, we contend that a means-based ethical evaluation of nudging equates proportionality with autonomy, manipulation, and transparency.

Proportionality has been a concept of law for an extensive period of time and is primarily used in assessing whether an intervention is constitutional, i.e., is the intervention in conflict with fundamental democratic values such as interference with individual rights. It is found in the Charter of Fundamental Rights of the European Union (Art. 52(1)) and plays a vital role in the adjudication of the European Court of Human Rights (Schweizer, 2016, p. 4). The level of scrutiny applied when assessing the proportionality of a measure differs greatly between jurisdictions; therefore, a discussion of the application of the principle of proportionality to nudging must remain at a very abstract level. Although this limits its usefulness greatly, we believe this to still be an extremely important criterion. For the purpose of an evaluation of nudging as a policy instrument, we will limit ourselves to the principle of proportionality in the context of fundamental rights. Any other applications of the principle of proportionality do not relate to nudging.

The principle of *proportionality* will be discussed here as a four-prong test that assesses (1) whether the measure serves a legitimate public interest, (2) whether a measure interfering with a right is suitable to achieve its objective, (3) whether it is necessary for that purpose and (4) whether it excessively burdens the individual compared with the benefits it aims to secure (Schweizer, 2015, p. 5).

1. A legitimate public interest can be any economic, cultural or social purpose. For instance, the protection of minors, minorities, socially disadvantaged persons, public health, the environment or the advancement of families or gender equality are all examples of legitimate public interests (Schweizer, 2015). The measure must be aimed at preventing social harm or a negative externality. Since we are purely discussing environmental nudges in this paper, we conclude that all nudges discussed here serve a legitimate public interest in the form of protecting the environment. We will, therefore, not examine the first prong in our evaluation of green nudges.
2. A measure is deemed suitable when it is, at least in principle, possible that it furthers the stated public goal (Schweizer, 2015, p. 15). It does not need to be the optimal means to further the public interest; it merely has to have the abstract potential to help achieve the goal. Few legislative measures ever fail this prong of the proportionality principle. The evidence that nudging works to some degree is overwhelming, and this is all that is required under the suitability test (we return to this point in our preliminary evaluation; see also the Annex for selected empirical examples of nudges in practice).
3. A means can be argued to be required (or necessary) when there is no other means available that is equally suitable to further the public interest (Schweizer, 2015, p. 16). When there are several equally effective means to achieve a given objective the state is required to choose the “mildest means”, i.e., the intervention least restrictive of personal liberties (van Aaken, 2015, p. 5). A measure can be described as less restrictive, for instance, if it interferes less intensely with a given fundamental right,

or affects fewer people. Schweizer (2015) argues that a nudge is generally less restrictive than a coercive measure (p. 16). He explains this by examining the facilitation of environmentally sustainable food in cafeterias. A measure could force the canteen operators to provide tongs rather than spoons for environmentally unsustainable food items. The coercive alternative of banning certain foods altogether, or increasing their price by taxation, is more coercive on the consumer, and equally or even more coercive on the operator.

4. The last prong of the proportionality test requires the undesired effects to be weighed against the importance of the public interest the means seek to achieve. Balancing requires “that the weight of the interference must not be disproportionate to the weight of the justifying reasons” (Schweizer, 2015, p. 19). The balancing test does not require that the measure is proportionate; it only requires that it is not disproportionate (Schweizer, 2015, p. 19). Considerable effort has been spent on structuring the balancing test and making it more predictable. For the purpose of this evaluation we will be using the balancing test by Alexy (2003), in which affected interests are grouped into three categories: ‘little’, ‘intermediate’ and ‘most’ worth protecting (pp. 780-789). In his analysis, Alexy (2003) argues that the highest category is reserved for conditional rights, such as freedom of religion. In the same way, Alexy (2003) categorises the public interest(s) the measure seeks to further with the protection of constitutional goods having the most weight. This category consists not only of the protection of the fundamental rights of others, but also the protection of constitutionally mandated public goods, such as the protection of the environment.

Any legal system that protects liberties is bound by law to use its ‘corrective power’ only proportionally and as prescribed by law. Therefore, measures that limit freedoms of any kind need to be justified, and when in doubt, one should always choose the mildest possible intervention (van Aaken, 2015, p. 4). Furthermore, it is important to note, that although coercive or mandated action can be legally challenged, persuasive or invisible measures are much harder to contest legally.

We have intentionally not included the level of intrusiveness, the fairness of the nudge and the critique that nudging can compromise citizen’s ability to form preferences in our evaluative framework. All nudges are to some degree intrusive, but the intrusiveness of a green nudge, we believe, is already encompassed by the criteria set out in our current framework. Concerning the latter two critiques, these are mentioned in the discussion section of our paper. We did not believe these to be necessary for our evaluative framework, as they are general ethical critiques as opposed to measureable ethical variances between different types of green nudges.

4.3 Methodological limitations and delimitations

This section serves to highlight potential limitations and delimitations to the current research.

4.3.1 Limitations

As the research conducted consists predominantly of a literature review, there is the risk of systematic bias within the material reviewed. Particularly studies from the behavioural sciences tend to lean towards a confirmatory bias of the appropriateness of green nudging – and fail to account for ethical concerns. Moreover, compared to the discussion on the effectiveness of nudges, ethical questions are not sufficiently addressed in policy circles. Due to this observation, we intentionally put an emphasis on ethics compared to the other two evaluative criteria and approach the ethical evaluation critically.

We also observe an overall positive bias in much of the available literature concerning the suitability of green nudging. One reason for this is that the topic is relatively new and very much centred around the book by Thaler and Sunstein (2008), who are strong proponents of the use of nudges in government. In addition, many scientists, who deal with the topic come from the private sector where nudging has already been implemented for years at varying levels of success. Lastly, there is a danger of overlooking the critical implications of nudging, especially with regards to ethical questions. The concept of libertarian paternalism in connection to nudging may create a positive view since it purports to preserve freedom of choice and make people happier by increasing welfare. Being aware of these positive biases in the literature helps us to actively seek out all positions in the debate and avoid falling in the trap of presenting only one perspective.

In studying the complexity of a policy process, researchers are restrained by cognitive presuppositions that cause them to recognise some aspects of the process and ignore others (Weible, 2014, p. 3). This cannot be avoided since illuminating certain aspects automatically means dismantling others. Furthermore, any research is confronted with limits in the scope of analysis, simply due to limited time resources, limited human capability as well as a limited amount of space to express the findings. Establishing a framework, as we have done here, can be one way to help mitigate such presuppositions and highlight the most important elements of the study (Weible, 2014, p. 3). Moreover, intentional limitations in scope can also favour more in-depth analysis of the chosen focal points – as long as the focus has been justified and reflected upon.

4.3.2 Delimitations

There are two preliminary assumptions underlying this paper, which we would like to stress. The first one is that we focus our evaluation of green nudging on the question of whether it is justified as a means. Thereby, as discussed in Section 3.4, we consider the ends to be justified, i.e., we take environmental protection as something worth being pursued. Thus, if evaluations suggest that green nudging is a legitimate and suitable policy instrument, the nudge would be justified as a means. We are aware of the fact that this assumption is debatable. However, we take this step for practical reasons. If we did not do so, we would

necessarily have to question environmental policymaking, or the value of environmental integrity in general before getting around to whether green nudges should be used to pursue such interests.

The current research does not use a complete, already existing evaluation framework because there is none that can fully cover all the aspects that are relevant when evaluating nudging as a policy instrument to tackle environmental problems. As we have shown in Section 3.2, nudging does not fit into the traditional environmental policy toolkit. Thus, the characteristics that are unique to nudging and new to the policy toolkit must be incorporated into the evaluation. Since these were not fully covered by other evaluation frameworks, we propose a framework that is tailored to the task of assessing green nudges by *inter alia* considering ethics with more weight than is currently standard.

The method used to approach the research question of this paper is an in-depth literature review. Therefore, we base our argumentation on existing literature pertaining to nudging – within and outside of the environmental field. Since the debate around the legitimacy of nudging as a policy tool has exploded over the last few years, there is a large amount of literature to be reviewed. Yet, not all the discussion takes place in an academic setting, as the topic has attracted great attention in the public sphere. Thus, in conducting our evaluation we investigated a mix of different kinds of literature on the topic – primary, secondary, popular, academic, working and discussion papers as well as technical reports and policy briefs by government institution – in order to cover the full research spectrum.

Finally, the cases of nudging we discuss come predominantly from the OECD countries. Future research could broaden the geographical scope.

5 A policy evaluation framework for nudging

Within the typology of nudges (Figure 2), two types clearly stand out in the literature: defaults and social norms. In the following, we use the framework outlined above to provide a preliminary evaluation of different cases of these two nudges in the environmental field. Unlike traditional evaluations, special attention is given here to the ethics criterion. While efficacy and effectiveness can employ quantitative, standardized metrics, the ethical criterion – in this case an innovative and integrative addition – requires a more qualitative approach.

5.1 Green Defaults

As described above, the attitude-behaviour gap describes how peoples' intentions and actions are not always consistency when it comes to the environment. When asked, for example, to hypothetically choose an electricity provider most people will name a green or environmentally friendly utility. However, actual green electricity consumption does not match reported intentions in countries (such as Germany) that facilitate choice of carrier (Pichert and Katsikopoulos, 2008). Numerous studies have shown that pre-set default options

have the power to promote environmentally conscious choices, including the selection of renewable electricity.

5.1.1 Efficacy

In the following, we focus on whether theories about the success of green defaults can be tested and validated in the context of methodological trials and experiments, namely efficacy studies. Subsequently, we provide an assessment of cases in which scholars such as Toft, Schuitema and Thøgersen (2014) or Pichert and Katsikopoulos (2008) (amongst others) have evaluated green defaults as efficacious or inefficacious.

In a series of experiments Toft et al. (2014) investigate whether defaults serve as strong policy tools to nudge people towards environmental ends. Consumers were also confronted with green defaults promoting smart grid technology, which is a technological prerequisite for the energy transition to renewables (Toft et al., 2014, pp. 119-121). The sample included Danish, Norwegian and Swiss electricity consumers. The framing of the question was manipulated: either as opt-in (“Tick the box below if you *accept* installation”), opt-out (“Tick the box below if you *do not accept* installation”) or a neutral active choice for one alternative or the other (Toft et al., 2014, p. 11). The aim of the analysis was not only to test the efficacy of the default in steering choices, but also to identify the type of default that maximises people’s willingness to adopt smart grid technology. The *theoretical basis* of this paper consists of the theory of nudging in regard to relevant environmental choice situations as well as the importance of cognitive biases when it comes to two types of instruments: information and default settings as part of nudging (Toft et al., 2014, pp. 4-5). The argumentation is that default nudges can be considered as one of the most powerful types of nudges for environmental purposes (Toft et al., p. 121). These theoretical assumptions rely mainly on research into peoples’ processing limitations and cognitive biases. In many cases individuals assume that a default setting signifies the better and safer option. In other cases, people’s laziness plays a large role.

As a *reliable method*, online surveys were conducted with the aim to test consumer behaviour in a constructed, laboratory setting. More specifically, the study used a “3 (types of benefits mentioned) x 2 (personal control mentioned or not) x 3 (framings) between-subjects design with random allocation to groups [...] to test the effects of different framings of the choice task while controlling for outcome and control information about smart grid technology” (p. 129). Participants were asked to first read a text about smart grid and later answer related questions. Here the default option was manipulated. This testing method shows high internal validity and strong evidence for the importance of default formulation. We assume that due to the random assignment of participants to control groups, the independent variable caused the observed outcome. The results of the online surveys in all three countries show that defaults framed as opt-outs lead to a significantly higher acceptance among participants, confirming the theoretical background that a strong default effect exists for environmental behaviour (Toft et al., 2014, p. 136). This leads to the conclusion that opt-out framing is a significantly effective way of nudging people in an ecological direction (Ölander

& Thøgersen, 2014, p. 349). Moreover, results show consistency with formerly conducted behavioural economic studies. For policymakers, this might be an informative result since default nudges cannot only be used by private energy companies but also on a public policy level. *Replication* of this hypothetical lab experiment was conducted under real-world conditions, namely in two field experiments in order to guarantee external and general validity. As we will describe thoroughly in the next Section 5.1.2, both studies also support the general hypothesis.

Another energy case in which the efficacy of a green default was tested is the choice between green and conventional electricity, evaluated by Pichert and Katsikopoulos (2008). The *theoretical basis* of this paper consists of psychological environmental studies that clearly suggest an attitude-behaviour gap on the part of the consumers when it comes to green energy consumption. Studies show that consumers are willing to pay a higher price for green energy but do not put this willingness into practice. Interesting for environmental policymakers is the assumption that people stick to defaults when trade-offs between saving money and environmental protection are made (Pichert & Katsikopoulos, 2008, p. 65). This is typically the situation when electricity consumers are confronted with the choice between a renewable or conventional electricity contract. Additionally, Pichert and Katsikopoulos (2008) conclude that due to further investigation regarding defaults, it becomes evident that people rarely switch the default options and choose the alternative. This might be because they have little knowledge about the product or programme or because they simply assume that the authorities who established the programme know better. As outlined in Section 2, these reactions are often due to cognitive biases (Pichert & Katsikopoulos, 2008, p. 65). Acting on the assumption that defaults play an essential role when decisions about money and the environment are being made, Pichert and Katsikopoulos (2008) sought to prove their hypothesis that “defaults affect the choice of electricity” (p. 65) in two laboratory experiments as well as in two natural experiments.

In our case, as a *reliable method*, only the first of the two laboratory experiments is relevant. Pichert and Katsikopoulos (2008) asked young adults to fill in a survey in which they had to imagine they were moving to a new city and thus obligated to choose a new electricity contract – between green and ‘grey’ (conventional energy). The information given to the participants involved a manipulation check to test whether participants understood that grey energy is significantly less environmentally friendly than green energy (Pichert & Katsikopoulos, 2008, p. 67). In total 225 individuals participated (63% students), each randomly assigned to one of three different experimental conditions – opt-out (green default), opt-in (grey default) and neutral (active choice). A power analysis was conducted in order to make sure that results of the selected sample would show at least 0.80 (reasonable level of statistical accuracy) “for a medium effect size at a significance level of 0.05” (Pichert & Katsikopoulos, 2008, p. 67). The lab experiment showed interesting results. For example, anticipated effort or laziness prevented many participants from choosing the non-default option. Often, distrust regarding the non-default option was visible and more information on the non-default

option was requested than on the default option. Only two per cent criticised the default as such. Thus, Pichert and Katsikopoulos's (2008) hypothesis that the default setting would influence respondents' choices was confirmed: The green energy alternative was chosen significantly more often when it was set as a default option (p. 68). Motivated by the fact that there is a significant gap between peoples' desire to use green energy and the percentage of customers that actually have a green energy contract, Pichert and Katsikopoulos (2008) confirmed their hypothesis using both laboratory and natural experiments. The *replication* of tests under real-world conditions as well as in the laboratory suggests high external validity.

So far, our evaluation can confirm strong internal validity for green defaults in the realm of environmentally friendly energy and the usage of smart grids, based on two extensive efficacy trials with similar positive outcomes. These examples, however, are not the only examples backing up the suitability of green nudges as an efficacious type of policy instrument. A different series of experiments, for example, evaluated the efficacy for default settings in regard to energy efficiency, particularly the use of energy-efficient light bulbs. In three experiments, Dinner, Johnson, Goldstein and Liu (2011) measured the impact of effort, implied endorsement and reference dependence on choice, adding an interesting new dimension to the discussion, namely the hypothesis that "defaults may act as instant endowments" (pp. 5-6). Identical to the studies described above, the experiment showed that green defaults mattered greatly and further supported prior experiments that had highlighted these three variables as responsible for the success of defaults (p. 2).

In reviewing the literature on the efficacy of green nudges, this research finds that defaults, as one type of green nudge, can be generally considered as efficacious. A common conclusion by all three groups of authors, who based their assumptions on sound psychological theories, is that default effects are particularly strong in a laboratory setting. Based on our literature review, green defaults show strong internal validity in many environmental realms. More effort, however, needs to be put into replication. Conducting a study several times within a laboratory setting, with different subject populations, adds external validity and indicates the generalisability of a phenomenon.

Nevertheless, an evaluation of the efficacy of a policy instrument or a type of policy is not enough. As previously mentioned, when testing consumer behaviour, researchers often highlight the attitude-behaviour gap – consumers' attitudes towards environmental ends may be positive but do not translate into action under real-world conditions. The next step is to test the effectiveness of a policy instrument under such conditions, paying close attention to the implications of the nudge in practice.

5.1.2 Effectiveness

Ölander and Thøgersen (2014) summarise the findings of the smart grid study, originally conducted by Toft et al. (2014) and introduced in the previous section. The authors draw the conclusion that defaults are a powerful tool to foster pro-environmental behaviour in

the form of green energy consumption. However, if we have a closer look at the details of the study and apply our effectiveness criteria, we find that one should consider a few additional aspects before applying defaults in environmental policies. First, it is worth mentioning a detail that Ölander and Thøgersen (2014) leave out in their evaluation of the study: It was not only conducted in a hypothetical way using online surveys about a potential installation of the technology, but also a field experiment was carried out in Denmark. Thus, not only the efficacy of defaults was tested, but also the effectiveness under real-world conditions (Toft et al., 2014, p. 115). This is important for our evaluation, since we established efficacy trials to be a prerequisite for any test of effectiveness.

Our criterion *practicability* is addressed by the fact that a field experiment was performed additionally to the hypothetical study in Denmark (as well as in Norway and Switzerland). The question is, though, whether practicability is not only addressed, but also sufficient to call defaults an effective type of green nudge. It could be argued that the study is practically relevant since the electricity company's query could be considered a real-world scenario. However, it is difficult for us to draw conclusions about the external validity of the study when so little information is given about the sample group. With no proven external validity, a green nudge cannot be called practicable, since it remains only valid in the hypothetical sphere.

The sample group in this particular study can be summarised as 140 Danish electricity consumers provided by a company that had conducted business with them in the past. However, Toft et al. (2014) point out that they received no socio-demographic data about the consumers (p. 119). This is a critical detail when we want to assess *generalisability*: If no data other than the addresses of the consumers are available, it is not possible to evaluate whether they represent the broader (Danish) population. Moreover, the number of participating persons may be too low to generalise the results. Although the authors claim the opposite with their remark that all three hypothetically tested countries showed similar results regarding the effectiveness of defaults, this holds only true for the efficacy of the default, but is not automatically proven in the real world (Toft et al., 2014, p. 120). If we not only take generalisability of the sample group into account, but also the setting and implementation of the study, we can conclude that the authors do not prove effectiveness of defaults for settings other than Danish households (as for example for workplaces) and neither do they mention any replications of the study in which other skilled staff than the policy developers replicate the same results proving effectiveness.

Two other important criteria for effectiveness, *longevity* and *side effects*, are hard to evaluate on the basis of the given information about the effectiveness trial by Toft et al. (2014). We can only hypothetically evaluate these aspects of the effectiveness of green defaults in this case: Replications of the study would be required to test the default's longevity. Since both green nudges and smart grid technology are young inventions, we need to leave the question of long-term effects to be answered in the future. The same is true for side effects:

Too little information is given to assess whether a green nudge in the form of a default has side effects.

Lastly, the question needs to be answered whether the green default tested, leads to the desired *outcome* and if its *impact* is relevant enough to deserve the effectiveness label. Evaluating the outcome of the study, it can be concluded that the desired outcome, i.e., a higher acceptance of smart grid technology adoption through the use of a default nudge was achieved. The results are statistically significant, suggesting that green policies could be successfully promoted if policymakers made people actively choose to opt-out of a default that is considered beneficial for the environment. In order to be able to evaluate the impact of the default better, it would be useful to have comparison studies arranged to shed light on the degree of effectiveness of green defaults compared to other environmental policy tools.

In the previous evaluation of the efficacy of defaults, this paper presented the case of green energy defaults. Pichert and Katsikopoulos (2008) emphasise that it is not easy to find real-world examples in which customers are offered green energy as a default. However, they were successful in providing instances of it in Southern Germany. It is important to note that strictly viewed, these natural experiments are not true experiments, but observations, since the researchers could not ensure an experimental control (Pichert & Katsikopoulos, 2008, pp. 65-69).

The first observation dealt with the town Schönau, Germany in which green electricity has been the default since a citizen initiative bought the local electricity grid and took over the management after the Chernobyl disaster in 1986, relying entirely on renewable energy. Interestingly, the town has not historically shown a high approval of the German Green Party, and its inhabitants tend to be more conservative. However, despite the contentiousness of the initial decision to overtake the incumbent utility, almost all the citizens (roughly 99%) stuck with the green energy default, even after the liberalisation of the electricity market in Germany in 1998. At this point people would have been able to switch back to cheaper grey energy (Pichert & Katsikopoulos, 2008, p. 66). In a second natural example of the effectiveness of green defaults, Pichert and Katsikopoulos (2008) describe the German energy company Energiedienst GmbH, which in 1999 broadened their energy offers, among other things with a green tariff. Energiedienst GmbH made the green energy option the default and it was even cheaper than the previous standard tariff. Besides the green default, the company offered an even cheaper grey option as well as a more expensive green alternative (financing newer facilities). Similar to the Schönau example, circa 94% of the customers kept the green default (Pichert & Katsikopoulos, 2008, p. 66).

To sum up, in these two cases, green energy defaults seemed to be highly effective because nearly all customers remained with the default choice. But what can we conclude when we use our effectiveness criteria with these examples? First, the *practicability* of these results is relatively high. Although the natural experiments are flawed from an empirical perspective since there is no experimental control (Pichert & Katsikopoulos, 2008, p. 66), they are

arguably very valuable for policymakers since they are proper real-world cases in which people were successfully nudged to make pro-environmental choices. Furthermore, the Energiedienst GmbH case provides evidence of 150,000 households of which 94% stayed with the green default, which not only gives *practical* support for default effectiveness, but also makes it generalisable due to the sample size.

The Schönau case, however, might be harder to prove regarding *generalisability*, since the sample size is relatively small (2,500 inhabitants) and “the unemployment rate is relatively low” (Pichert & Katsikopoulos, 2008, p. 66). Thus, the case of the small idyllic Southern German town might be difficult to generalise to a rather heterogeneous larger city like Berlin. As in the smart grid case, the setting is limited to private households and not generalisable to other settings, but the implementation contrasts with the other case independent from the researcher or policy developer, which naturally happened through the non-experimental/non-empirical setting of the German examples.

Whereas possible *side effects* cannot be evaluated with the information provided (side effects are not mentioned in the study), the German cases serve as interesting ones with respect to *longevity*. Especially Schönau provides effectiveness evidence: According to Pichert and Katsikopoulos (2008), the default was still highly effective after it had been set for eight years (1669 of 1683 electricity meters were still supplied with green energy in 2006). It would be interesting to see the numbers today, a decade later, but even with the data from 2006 it can be assumed that the green default had a long-term effect. Needless to say after mentioning the percentage of customers that stick with the green default, the resulting *outcome* was in line with Pichert and Katsikopoulos’s (2008) intention, namely that using green energy as a default option leads to a significantly higher share of green energy among consumers. Similar to the smart grid cases, the *impact* should be evaluated in comparison to other environmental policy instruments and further case studies in order to draw a safe conclusion for the effectiveness evaluation.

In conclusion, our preliminary evaluation of green default effectiveness suggests that Thaler and Sunstein’s ‘marketing’ of defaults as an effective nudge, shared by many scholars (Toft et al., 2014; Pichert & Katsikopoulos, 2008), can to a certain extent be confirmed by our evaluation criteria. As the cases in the field of green energy consumption show, green defaults seem to produce the desired outcomes and have proven, where data was available, to be effective in the long term. Moreover, they are effective with regard to their practicality, as the selected examples have been tested both under real-world and laboratory conditions. Extensive further evidence for the long-term effectiveness as well as the practicality of green defaults can be drawn from the summary of relevant empirical studies found in the Annex to this report. However, more research is required in regards to side effects, impact and generalisability. This is mainly due to the fact that green defaults, like green nudges in general, are a relatively new subject of research.

5.1.3 Ethics

Autonomy

When examining whether a particular green default nudge infringes upon a person's autonomy, and to what degree, we must scrutinise whether external forces distort that person's ability to make decisions based on their own reasons and motives. In the literature, there are arguments for and against the infringement of autonomy by green default nudges. Smith, Goldstein and Johnson (2013) maintain that default nudges are superior to active choosing on autonomy grounds, arguing that it is an insult to an individual's autonomy to force him or her to choose (p. 16). Further reasoning suggests that people who would like to choose, are not deprived of that choice by a default rule as they are free to reject the default.

Sunstein (2013), on the other hand, has argued that green default nudges might well intrude on autonomy even though freedom of choice is preserved. For many green nudges, including default nudges, the causal mechanism through which choice architecture is intended to work "[...] deliberately seek[s] to by-pass the individual's rational decision-making processes in order to channel behaviour in the direction preferred by the choice architect" (Yeung, 2012, p. 135). Default schemes from which individuals may opt out of easily and cheaply, seek to exploit status quo bias and inertia to nudge individuals into adopting the default preference. When examining green default nudges in this way, they entail a subtle form of manipulation by taking advantage of the human tendency to act unreflectively and, to that extent, are inconsistent with demonstrating respect for individual autonomy (Yeung, 2011, p. 136). Default nudges purposefully seek to exploit the tendency of individuals to choose passively and unreflectively (Yeung, 2011, p. 137).

Smith et al. (2013) believe that there is a good chance, due to the force of inertia, that people might not reject a harmful default (p. 161). If so, there is arguably an intrusion on a person's autonomy, as they will end up with outcomes that they did not specifically select. Three basic psychological factors have been suggested to be causally responsible for the behavioural impact of defaults: inertia, loss aversion and implicit recommendations (Sunstein & Reisch, 2013, pp. 140-144; Smith et al., 2013; Schubert, 2016, p. 18). An extensive amount of literature claims that these three processes are all external forces: distorting a person's ability to make decisions based on their own reasons and motives. We must, therefore, conclude that green default nudges infringe, at least to some degree, on a person's autonomy.

Manipulation and transparency

Evaluating whether a green default nudge is manipulative, similar to evaluating its effect on autonomy, is a process not simplified by examining the literature. Thaler and Sunstein (2008), for instance, dismiss the notion that nudging is a manipulation of choices, by arguing that, because green default nudges work without limiting the original set of choices, citizens remain free to choose otherwise (p. 82 and p. 239). Others, such as Binder (2014), claim that since default rules work particularly well on people who do not reflect on the decision

to be made – and are thus not even aware that they are opting out – the nudge is manipulative in nature (p. 1118).

Recall our definition of manipulation: the attempt to influence people subconsciously, in a way that undermines their capacity for conscious choice. We identified a framework used by Hansen and Jespersen (2013) to outline the level of manipulation of various nudges. In this framework, all nudges are categorised as either transparent or non-transparent, as well as whether they are Type 1 or Type 2, i.e., similar to the passive/activating distinction in our typology (refer to Figure 2).¹¹ We will be closely using Hansen and Jespersen’s (2013) framework to make our preliminary evaluation of the level of manipulation and transparency of green default nudges.

Hansen and Jespersen (2013), interestingly categorise green default nudges as both transparent and non-transparent Type 1 nudges. A green default nudge falling into the category of a transparent nudge could, for example, be setting the default printer option to double-sided printing, as opposed to single-sided, to reduce paper consumption. Rutgers University adopted this double-sided printing default, resulting in a 44% reduction of paper consumed in the first three years (Sunstein & Reisch, 2013, p. 399). Green default nudges that fall under the category of transparent Type 1 nudges do not try to influence citizens’ choices. Rather, they involve influencing automatic behaviours and the consequences thereof in a transparent way. Hansen and Jespersen (2013) claim that this type of influence is difficult, if not impossible, to avoid as it activates instinctive responses (p. 24). Although, they conclude that these types of green default nudges do in fact manipulate, it is noted that this is in the sense of ‘technical’ manipulation, not psychological manipulation (Hansen & Jespersen, 2013, p. 24).¹²

Although citizens are technically free to ignore the default and choose to change their own preferences should they wish, this freedom is largely theoretical since exposure to a green default nudge will affect preference selection through automatic behaviour (see Annex for multiple examples). While the transparent element of this type of green nudge makes it possible for citizens to recognise the intention behind and means by which their behaviour

¹¹ As discussed in Section 3.3.1: While both categories of nudges aim at influencing automatic modes of thinking, activating nudges aim at influencing the attention and premises of reflective thinking (i.e. choices), via influencing the automatic system. Passive nudges on the other hand, aim at influencing the behaviour maintained by automatic thinking, or consequences thereof without involving reflective thinking (Hansen & Jespersen, 2013, p. 14).

¹² Hansen and Jespersen (2013) define *technical manipulation* as “the intentional manipulation of a straightforward cause-and-effect relationship” (p. 19). *Psychological manipulation*, on the other hand, is defined as “manipulation in the sense of intending to change the perception, choices or behaviour of others through underhanded deceptive, or even abusive tactics” (p. 18). They further argue that nudging usually only changes frequencies and thus the effect is probabilistic rather than deterministic (p. 19).

is influenced, it does not easily allow them to avoid this. The use of nudges such as these should be generally regarded as acceptable; they allow citizens to easily dispute their influence within the democratic process, and assign proper responsibilities to public policymakers.

Non-transparent Type 1 nudges also include green default nudges. An example of a green nudge from this category is the opt-out rather than opt-in smart grid trial conducted by Ölander and Thøgersen (2014) and discussed in the previous section. The authors report on a study that examined consumers' willingness to participate in a smart grid trial where their household's consumption could be automatically reduced at peak electricity demand periods. The opt-in to the trial (choose to participate) rendered a 60% participation rate, whereas the opt-out option (choose not to participate) rendered a participation rate of almost 80% (Ölander & Thøgersen, 2014, p. 351).

Green default nudges of this manner cause behavioural change without engaging the reflective system and in such a way that makes recognising the nudge extremely difficult. Therefore, we conclude that nudges of this nature influence behaviour in a non-transparent way, and as such, their application constitutes the use of both technical and psychological manipulation. Citizens are in general only capable of avoiding their effects as a matter of principle, avoiding it in a complex everyday setting seems much more difficult, if not impossible.

Proportionality

When examining green default nudges in regard to the principle of proportionality, we must evaluate whether this type of nudge passes the four-prong test. As mentioned in our evaluative framework (Section 4.2.3), we assume the ends of all green nudges to be justified; therefore, we will not discuss the first prong of the proportionality evaluation any further. The literature on green nudges provides many examples of where green default nudges work, at least to some degree, and since this is all that is required to fulfil the second prong of suitability, green default nudges fulfil the second prong also.¹³

Exploring the third prong is where we run into difficulty with our preliminary evaluation. To sufficiently answer whether green default nudges are or are not the only suitable measure to further the public interest, an in-depth evaluation would need to be done of all other possible measures. As we have neither the time, nor the available resources to conduct such an evaluation, we will conclude that since the likely alternative to default nudges is active choice, green default nudges are the most suitable measure. We conclude this on the basis that green default nudges have shown to further the public interest significantly more than active choices. One example of this can be seen in a pair of studies conducted by Pichert

¹³ Refer to Ölander and Thøgersen (2014), Pichert and Katsikopoulos (2008), Thaler and Sunstein (2008).

and Katsikopoulos (2008) in which 94-99% of consumers stayed with the green electricity default rather than switching to a cheaper but fossil-based electricity supply.

The final prong produces the largest problem for the current paper. This part of the preliminary evaluation requires us to weigh the undesired effects, predominantly the ethical criticisms discussed earlier, against the importance of the public interest the green default nudge is seeking to achieve. Green default nudges are always seeking to reduce negative environmental externalities, such as reducing the effects of climate change through small scale acts of mitigation or adaptation. Due to the tremendous importance climate change has, and will increase to have, on human security it is likely that there are few cases in which ethical criticisms would outweigh a green default nudge in theory. Of course, each individual green default nudge, prior to being implemented, would still need to be evaluated independently to determine the ethical implications of its use.

5.2 Green Social Norms

Most humans are strongly and regularly affected by social norms that guide their actions directly and indirectly in many meaningful ways. A normal propensity for most humans is to learn social norms, which can be understood as the standard behaviour and shared understanding of certain actions based on a common belief system. Recall that social norms can be categorised as descriptive, comparative or injunctive. Descriptive and comparative norms refer to neutral information about behaviour in society, whereas injunctive norms attach moral weight to certain forms of behaviour.

As with littering, many social norms already define our behaviour towards environmental issues. However, global environmental problems, such as climate change, call for a fundamental and rapid change in social norms to guarantee lasting progress. On this basis, we investigate the role of social norm nudges in the context of environmental policy goals.

5.2.1 Efficacy

One of the chapters in Thaler and Sunstein's *Nudge: Improving Decisions about Health, Wealth and Happiness* deals extensively with social norms. In many ways, humans are highly influenced by the behaviour of other humans; sometimes to the extent that we do not need to have contact with other humans in order to stick to certain rules and norms. In such cases we are guided by social norms. In order to understand 'following the herd' behaviour we must revisit the difference between *Econs* and *Humans*: "Humans [...] are frequently nudged by other Humans. Sometimes massive social changes, in markets and politics alike, start with a small social nudge" (Thaler & Sunstein, 2008, p. 64). Self-interested *Econs* are not swayed by the actions or beliefs of their peers. This sounds promising for environmental policymakers, since the social aspect of certain nudges might help to effectively foster pro-environmental behaviour.

Ölander and Thøgersen (2014) state that the social character of humans certainly is a source of environmental problems, but that it could also serve as a solution. A study on energy

conservation through ‘product-integrated feedback’ by McCalley and Midden (2002), is a good example of the power of green social norms and their efficacy. In the following, we take a closer look at examples regarding social norms and again apply our efficacy criteria. The question is whether the scholarly consensus about the power of green social norm nudges can be confirmed from a laboratory point of view.

Generally, theories on social norms are mainly based on why we follow norms and how we learn them. Particularly in the environmental realm, some differentiate between two different types of social norms: “social norms of conformity or cooperation and pro-environment social norms” (Kinzig et al., 2013, p. 170). As norms of conformity or cooperation are far more universal and do not interfere with other interests or values, they may be more suitable for reaching environmental goals.

Many efficacy studies have been conducted in regard to social norms and alcohol consumption. Employing RCTs, Lewis and Neighbors (2006) found that personalised feedback could help reduce units of alcohol consumed per average occasion. Due to a lack of laboratory studies on the efficacy of social norms, however, we were forced to limit our analysis to one paper. McCalley and Midden (2002) evaluated the effectiveness of feedback intervention and commitments in a laboratory setting, asking participants about their energy conservation behaviour.

The *theoretical basis* of this paper is very extensive and is mainly comprised of psychological studies (including former efficacy studies) about the relationship between feedback and goals, particularly ‘feedback intervention theory’ (McCalley & Midden, 2002, p. 594). Based on these studies, the authors assume that goal-setting is needed as a reference in order to understand the effects of feedback. Following this logic, if a task is comprised of goal-setting and goal acceptance, feedback effects should be automatically improved (McCalley & Midden, 2002, p. 594). The study itself focuses on “[...] the integration of feedback into a system designed for a single task [one household appliance]” (McCalley & Midden, 2002, p. 593).

The *reliable method* of the study consisted of a computerised task, representing ten washing machine trials within a graphic representation. Participants included one hundred residents from the Dutch city of Eindhoven who fulfilled the criterion of washing once a week. First, they received questions concerning demographic data and participated in a game in order to test social orientation. In the following sessions, four feedback groups were used in order to test “(1) feedback with no-goal manipulation, (2) feedback with a self-set goal, and (3) feedback with an experimenter assigned goal [as well as] a fourth group [...] tested under a no-feedback-no-goal condition and served as the baseline control” (McCalley & Midden, 2002, pp. 595-596). Hypotheses were tested with a 3x2 GLM factorial design (McCalley &

Midden, 2002, p. 598).¹⁴ Interestingly, evaluations of the trial revealed that feedback is clearly related to goal-setting. Both groups, self-set and assigned, significantly reduced their energy consumption (21.9% and 19.5%) (McCalley & Midden, 2002, p. 599).

This single study supports the efficacy and internal validity of green social norms, specifically commitment or goal-setting, based on one example in the environmental realm. Moreover, the study is a *replication* of a formerly conducted study by Becker (1978) and therefore equally serves to support the generalisability of the outcome. Notably, the majority of social norm experiments (particularly those relevant to the environment) are field studies and thus covered in the following section.

5.2.2 Effectiveness

There is ample evidence to back up Thaler and Sunstein's (2008) claim about the high level of effectiveness of social norms nudges. For instance, van der Heijden and Kusters (2015) as well as Cialdini et al. (1990) mention the remarkable effect of normative values with respect to public littering. Furthermore, Cialdini et al. (2006) examine the problem of theft of petrified wood in a national park in Arizona, which was to a great extent solved by the use of a green social norm nudge. However, as stated above, the most discussed cases in which the effectiveness of green social norm nudges has been evaluated are: reusing towels in hotels because other guests do so (Cialdini, 2005; Cialdini & Goldstein, 2004; Goldstein, Cialdini & Griskevicius, 2008; Schultz, Khazian & Zaleski, 2008) or saving energy at home because neighbours do (Graffeo et al., 2015; Schultz et al., 2007; Costa & Kahn, 2010; Allcott, 2011; Ölander & Thøgersen, 2014).

As mentioned above, the idea to nudge hotel guests to reuse their towels as an energy saving measure has been discussed extensively in the literature. Goldstein et al. (2008), for one, conclude that the most effective way to make people reuse their towels is not by appealing to people's moral drive to help save the environment, but instead with a social norm that states, "the majority of guests in this room reuse their towels" (p. 1). In their first field experiment, Goldstein et al. (2008) compared the effects of signs in hotel rooms that use environmental protection messaging with signs that employ green social norms (stating that most other guests reuse their towels). The result confirmed their hypothesis – the green social norm was significantly more successful in promoting conservation behaviour. If we evaluate this experiment on the basis of our effectiveness criteria, can we conclude that green social norms play an essential role in effectively changing people's behaviour?

With respect to *practicability*, it is worth mentioning that Goldstein et al. (2008) did not inform the hotel guests about their participation in the experiment. Thus, the conditions can be considered useful to test practicability. Since the effect of social norms is tested in

¹⁴ General Linear Design (GLM) can be seen as an extension of multiple regression (for a discussion see Cnaan, Laird & Slasor, 1997, pp. 2349-2351).

contrast to the environmental protection condition, which is the conventional attempt in hotels to save energy, we have a control group that the social norm nudge can be tested against (Goldstein et al., 2008, p. 2). At least in our preliminary evaluation, the criterion practicability seems to be addressed. Related to practicability is generalisability, because effectiveness as defined by relevance in the real world is closely connected to whether the green nudge can be applied to broader conditions.

Regarding the *generalisability* to a broader population, it can be summarised that the study was conducted with about 1000 instances of towel (re)use in a hotel of medium size and medium price in the US. This could be regarded as a suitable setting to generalise the results. However, the setting might be more difficult to apply to other environments, as the stay in a hotel, at least for the average person, cannot be generalised as an everyday experience. Admittedly, it does not make sense to test this particular nudge in households, as people normally reuse their towels multiple times anyway. Still, one could extend the study in a similar way in a public setting by slightly changing the set-up. For instance, it could be tested whether norm-based, comparable messages work in schools. The amount of paper towels could probably be reduced if a sign said that other students only use one per time. Implementation apparently does not depend on the policy developers, as the nudge can be easily operated by anyone (see Goldstein et al., 2008, p. 3 for a description of how the hotel staff were trained).

Whereas the evaluation of practicability and generalisability is a relatively straightforward procedure due to the provided data about the sample group and other conditions of the effectiveness test, it is more difficult to assess *longevity* and *side effects*. Whether green social norms in terms of towel usage in hotels are effective on a long-term basis would be an interesting but challenging focus of further research. What the study by Goldstein et al. (2008) does not address is whether the towel nudge remains effective when hotel guests are confronted with it multiple times during further hotel visits. Therefore, longevity would have to be considered to draw valid conclusions about the effectiveness of green social norms. The same is true for side effects: The fact that Goldstein et al. (2008) do not address any does not mean that they do not exist.

To conclude our evaluation criteria, some thoughts follow about the questions of *outcome* and *impact*. The outcome could certainly be evaluated as a success. The researchers' intention was to save energy through a reduction of new towel usage in a hotel, which is reflected in the result of the test: Compared to about 35% of guests reusing their towels when a sign is placed that asks for environmental protection, roughly 44% do so when the social comparison is activated in their mind (Goldstein et al., 2008, p. 3). Little information is given with respect to the impact of the nudge other than an implicit connection between water conservation and towel reuse, and regardless further replications of the experiment are needed to determine the environmental effect this green social norm. However, as mentioned before, the nudge's practicability in a real-world situation suggests that social norms are theoretically effective in terms of impact.

Inspired by the success of social norms marketing, Schultz et al. (2007) conducted field research on green social norms and energy conservation with the intention of applying these marketing strategies to the environmental field. Since they specifically address the fact that green social norm can lead to so-called boomerang effects (i.e. undesired negative side effects, see Schultz et al., 2007, p. 429), this study is particularly interesting for our purposes.

In summarising the experiment, which we already briefly touched on in the Introduction and Section 3.3 of this paper, it should be mentioned that Schultz et al. (2007) tested the effect of letters sent to 290 households in California. In one of the sample groups, the message contained information about the households' energy consumption in the previous week and the average energy consumption of the neighbours. In the other group, a happy or sad smiley was added, dependent on whether the household had comparatively high or low energy use (Schultz et al., 2007, pp. 430-431). The findings confirm Schultz et al.'s (2007) hypothesis that energy consumption can be significantly decreased with help of green social norm nudges (pp. 431-432).

Our evaluation of the energy nudge's effectiveness starts with its *practicability*. Contrary to the towel experiment, the participants were aware of their participation in the study. Arguably, this could have had an impact on their behaviour via demand characteristics such as the social desirability effect, as they might want to perform well when being observed in their energy use. The aspect of the division of sample groups in two, receiving a message with or without a smiley, could have been added as another element to improve the overall practicability of the study: another sample group that receives a letter only with information on energy consumption and consequences for the environment (similar as in the towel case), so that the effect of social norms can be compared.

The *generalisability* of the results could be improved upon as well. In order to have a higher validity about the effectiveness of social norms, more participants as well as information about socio-demographic data would have been helpful. If the sample only consists of well-educated people, this can clearly have an impact on their responsiveness to the letter and it would be harder to generalise the findings. Furthermore, the study could be extended to different settings than households, for example testing the effects of the nudge in state institutions in which the manager does not, to the same extent as a private household, benefit from the money saved through less energy consumption.

A much more positive conclusion can be drawn from the testing of *longevity*. Schultz et al. (2007) address short- and long-term changes. Short-term changes resulting from the letters sent to households could clearly be observed: Households that had higher energy consumption than the average neighbour significantly reduced it and energy consumers below the average did not increase energy use when the smiley was added to the descriptive normative message (for an overview of effect sizes, see Annex). Long-term effects are almost identical in their outcome (Schultz et al., 2007, pp. 431-432). This could be an important and assuring factor for environmental policymakers if there was not one noteworthy detail: The authors only continued the study for four more weeks (p. 433).

As to possible *side effects* of the evaluated social norm nudge we have already touched upon the fact that the smiley added to the descriptive normative message plays an essential role to avoid the increase of energy consumption in households consuming less energy than the average neighbour (Schultz et al., 2007, p. 432). Thus, it needs to be mentioned that there would certainly be side effects if the researchers had not added the injunctive message with the smiley as an extra inductive social norm nudge. The creation of two sample groups that only differ in the smiley aspect can be considered very useful and novel. Hence, we conclude that side effects are well addressed in the effectiveness trial.

To finish our effectiveness evaluation of the presented nudge, we propose a brief assessment of its *outcome* and *impact*. Generally, the outcome can be evaluated positively. In each case, the researchers were able to show that social norms can indeed be an adequate tool to improve the environment by saving energy or reducing towel usage. The predicted boomerang effect that occurred when people realised they were using comparatively little energy could be counteracted successfully with the addition of an injunctive social norm nudge. The impact of the injunctive message is remarkable and adds a valuable new dimension to the effectiveness of social norm nudges.

The energy conservation intervention described above has been pilot-tested at scale in California by a company called OPOWER. The company has partnered with utilities in order to mail the social norm energy reports to thousands of homes across the United States (US). The OPOWER initiative has been shown to reduce monthly energy consumption by two per cent below a baseline (Allcott, 2011).¹⁵ If scaled nationally in the US, the green social norm intervention described above alone would decrease carbon dioxide emissions from the electricity sector by one per cent, avoiding 12.7 million metric tonnes of emissions (Allcott & Mullainathan, 2010, p. 1205). Thus, green social norms are already at the pilot-testing stage in terms of effectiveness, suggesting scalable and significant results.

Overall, it can be concluded that social norm nudges have, with the exception of a few weaknesses in single sub-criteria, which we have just addressed, a relatively high potential to be considered effective environmental policy tools. The flaws that were identified can to a great extent be explained by the fact that the two nudges evaluated are relatively new and therefore need further longitudinal research dedicated to longevity and impact. This is similar to our conclusion about the effectiveness of defaults. A summary of further empirical evidence for effective social norm nudges has been incorporated in the Annex to this report.

¹⁵ The study in question conducted a field experiment involving a sample of 80,000 households in Minnesota that were randomly assigned into control and treatment groups.

5.2.3 Ethics

Autonomy

Examining whether green social norm nudges infringe upon a person's autonomy, and again to what degree, requires the scrutiny of external influences and whether these distort a person's ability to make decisions based on personal reasons and motives. Schubert (2016) explains that one can "[...] communicate social norms about approved or disapproved choices, thereby harnessing people's desire to conform to social expectations, to draw 'moral utility' from that fact [...], and to engage in conditional cooperation [...]; this can be achieved by offering peer comparisons [...]" (p. 11).

Although Hacker (2016) makes a convincing argument in which he states that because both the content of the information and the behavioural tool is obvious and transparent, the autonomy of citizens is respected (p. 20). Others, such as Nolan et al. (2008), argue that because the awareness of what other people do has such a powerful influence on people's behaviour, a social norm nudge cannot be classified so easily as such (p. 921). The insight that human beings are inherently social animals may in fact be behavioural economics' key contribution to the social sciences (Gowdy, 2008, p. 642).

There are a variety of ways to harness this basic human characteristic in the interest of promoting green behaviour. Eco-labels, for example, are a key instrument in providing consumers with information about product characteristics, deemed by some essential for environmentally responsible behaviour. Impacting behaviour by increasing the *salience* of certain product characteristics, therefore, making consumers more aware of them. Félonneau and Becker (2008) argue that eco-labelling also serves to confer a certain social value on these characteristics, at least insofar as pro-environmental behaviour is in fact socially approved in the given socio-cultural context (pp. 33-35). Of course, eco-labels can also be used to convey social norms, thereby activating herd behaviour. This can also be said about providing energy users with simple feedback regarding their current energy consumption, such as with smart meters (Carroll, Lyons & Denny, 2014; Joachain & Klopfert, 2014).

One of the most successful applications of green nudges to reduce littering is the 'Don't mess with Texas' social advertising campaign, initiated in 1986 (Thaler & Sunstein, 2008, p. 64). It is estimated that this campaign reduced littering on Texas highways by about 71% between 1986 and 1990 (Mols, Haslam, Jetten & Steffens, 2015, p. 93). The slogan appears to target people's sense of community pride, framing littering as a kind of behaviour Texans find unacceptable (see also Grasmick, Bursik & Kinsey, 1991). Mols et al. (2015) argue that the effectiveness and sustainability of this particular nudge are largely due to its impact on people's self-understanding or *social identity*, making people internalise a specific social norm.

Nagatsu (2015), discussing the 'Don't Mess With Texas' campaign, claims that the objection based on infringed autonomy does not carry any weight against social nudges that exploit

these mechanisms (p. 488). The foundation for Nagatsu's (2015) argument lies in the differentiation between practical and theoretical reasoning, whereby social norm engineering operates through practical reasoning and is thus unproblematic (p. 9).¹⁶ Further emphasising that since social norm engineering operates through practical reasoning, autonomy is not infringed. Another account of green social norm nudges refers to the frame-based mechanism. Social nudges may induce people to shift from *I-frame* to *we-frame* in social dilemmas, thereby increasing pro-social, group-oriented behaviour. The key question, when considering whether autonomy is infringed upon, is whether a frame-shift is responsive to practical reasoning. If it is not, nudging people to adopt one frame rather than another does not appear to respect their autonomy.

We would, based on the literature, have to conclude that green social norm nudges that impact people's social identity, such as eco-labelling or the 'Don't Mess With Texas' campaign, do not infringe upon people's autonomy. We would, however, have to agree with Nagatsu (2015), that a frame-shift increasing group-oriented behaviour does not respect autonomy and is therefore not considered ethical under the criteria of autonomy employed by this framework.

Manipulation and transparency

For the evaluation of the level of manipulation and transparency of green social norm nudges, we will again use the framework established by Hansen and Jespersen (2013) as described in the evaluative framework. Green social norm nudges differ from green default nudges significantly here, as these are categorised in our typology as activating (Type 2) nudges. Although both nudging types aim at influencing automatic modes of thinking, Type 2 nudges are aimed at influencing the attention and premise of – and hence the behaviour anchored in – reflective thinking (i.e. choice) (Hansen & Jespersen, 2013, p. 14). These nudges are influencing behaviours best characterised as actions, the results of deliberation, judgement, and choice. Green social norm nudges are further categorised as transparent nudges.

A transparent activating nudge “[...] engages the reflective system in a way that makes it easy for the citizen to reconstruct the intentions and means by which behaviour change is pursued” (Hansen & Jespersen, 2013, p. 20). An example of a green social norm nudge is the use of social salience. This could be, for instance, the use of eco-labels, which convey social norms, thereby activating herd behaviour. Another example that falls into this category is the elicitation of descriptive norms with a clear messenger, for instance, a pamphlet showing

¹⁶ Practical reasoning, in contrast to theoretical reasoning, refers to a process whereby reason is used to arrive at a given decision. Interestingly, Nagatsu (2015) does not consider whether an action was conscious or unconscious as a relevant criterion for whether that action was delivered through practical reasoning or not: Everyday choices – frequently made unconsciously – are still considered to be authored by us (Nagatsu, 2015, p. 9).

the energy consumption compared to that of neighbours, or the ‘Don’t Mess With Texas’ campaign, discussed in the previous section.

Hansen and Jespersen (2013) conclude that although it may be obvious that this type of nudge influences behaviour anchored in reflective thinking, they are not aimed at doing so by means of psychological manipulation.¹⁷ Green social norm nudges of this type aim at promoting decision making in ways that are transparent to the agents influenced. They generally work by prompting choices consistent with the reflected preferences of citizens by making features, actions, preferences, and/or consequences salient, or by providing feedback (Hansen & Jespersen, 2013, p. 24). These types of nudges “[...] actually allow citizens to be nudged, to change their actions and behaviour in a predictable way, *while simultaneously leaving them free to choose otherwise – not just as a matter of principle, but also in practice*” (Hansen & Jespersen, 2013, p. 24).

Proportionality

Evaluating green social norm nudges in regard to the principle of proportionality requires assessing whether the guidelines of the four-prong test is met. In this paper we assume, due to reasons laid out in Section 3.4, that the ends of green nudges (i.e. protecting and upholding environmental integrity) are justified. Additionally, there are plenty of examples described in the literature where green social norm nudges work, at least to some degree.¹⁸ Since this is all that is required to fulfil the second prong, we conclude that green social norm nudges fulfil the second criteria as well. We can move on quite quickly to the third (necessity of measure) and fourth (weighing of effects) prongs of the principle of proportionality.

Exploring the third prong (whether it is necessary) is difficult to address adequately within the scope of this paper. To sufficiently answer whether green social norm nudges are or are not the only suitable measure to further the public interest, an in-depth evaluation of all possible measures would need to be conducted. We would like to, however, draw a conclusion based on the extensive literature we have reviewed. Based on the conclusion that green social norm nudges are both transparent and non-manipulative, we further conclude that nudges of this kind are the least restrictive intervention and, therefore, pass the third prong of the principle of proportionality test.

The last prong of the proportionality test requires us to weigh the undesired effects against the importance of the public interest. This, again, poses the largest problem for us. Regarding the evaluation of green social norm nudges, the undesired effects are the ethical criticisms discussed earlier, weighed against the public interest, in this case the preservation of

¹⁷ See Section 3.3.1

¹⁸ e.g. Kallbekken and Sælen (2012), Thaler and Sunstein (2008), Mols et al. (2015), Goldstein et al. (2008), see also the Annex

the environment. At the risk of being predictable, we must draw the same conclusion as in the previous evaluation of green social norms. Climate change and other global environmental problems will most likely negatively affect billions of people worldwide. Mitigation and adaptation actions, all the way down to the level of individual behaviour, are of a grave importance, and the involvement of citizens in policy implementation is crucial. In conclusion, it seems probable that – similar to green defaults – there are few circumstances under which ethical criticisms prohibit the use of a green social norm nudge. Of course, each green social norm nudge, before implementation, would need to be evaluated on an individual basis – considering contextual considerations.

6 Discussion

With two preliminary evaluations, we demonstrate strong support for both the efficacy as well as effectiveness of green defaults and green social norms. These findings fall in line with the rather positive outlook in the literature on nudging in general and in specific environmental contexts. However, we contend that the complicated ethical dimensions of green nudging call for a more integrative and comprehensive evaluation. Not surprisingly, researchers that take a closer look at ethical questions regarding nudging often position themselves as more cautious and do not necessarily share the positive assumptions concerning the new policy tool.

In terms of efficacy, existing applied laboratory research is based on an extensive theoretical foundation that includes dual process theory and the heuristics and biases approach to bounded rationality, both described at the beginning of this paper. Green defaults, for one, have undergone thorough laboratory experimentation but studies conducted in different environmental fields support the efficacy of both types of nudges. There seems to be widespread acceptance in the literature that the methods used to test these and other nudges are reliable. Indeed, in most of the cases evaluated here, control trials, including randomisation, were used to ensure internal validity in a laboratory setting.

The criterion of efficacy seems to be sufficiently addressed for green defaults. Green social norms, however, lack reliable laboratory experimentation in the environmental realm. Researchers such as Goldstein et al. (2008) who conduct field studies on social norms back up their hypotheses with references to laboratory trials, but these are often not environmentally related. Another major limitation in terms of the efficacy of both nudges is replication. Particularly for defaults, a lack of replication and small sample sizes both weaken the generalisability of findings to different contexts. When it comes to social norms, replication has been pursued to a far greater extent in field settings, including a handful of scaled pilot trials. Indeed, given how conducive social norm manipulations are to field experimentation, the literature seems to skip laboratory examination altogether. In sum, our evaluation suggests that research on social norms focuses much more on effectiveness trials, while default studies show a broader range of efficacy experiments and often use field experiments to bolster the external validity and generalisability of laboratory findings.

Efficacy does not guarantee effectiveness. Nevertheless, for both nudges most of our effectiveness criteria were sufficiently addressed in real-world, field experiments. Especially the sub-criteria of practicability and outcome play a central role in the examined cases. Both nudges evaluated resulted in significant changes in observed behaviour and the conditions of the experiments reflected real-world scenarios to a considerable extent. Furthermore, when side effects were identified in one case, researchers addressed them using a counteracting measure to achieve the desired outcome (e.g. the boomerang effect in the energy comparison between neighbours in Schultz et al., 2007).

However, our evaluation was not entirely positive. Notably, both nudge types show some weaknesses in terms of generalisability and studies tend to neglect longevity and impact. In terms of generalisability, it would be desirable to have more information and socio-demographic data about the sample. Longevity is not sufficiently addressed due to the fact that green nudges are a new phenomenon; there is not a longstanding research programme assessing them in the field. Only one experiment in our evaluation attempted to get a (preliminary) impression about longevity of effect for the (arguably unsatisfactory) duration of four weeks.¹⁹ Lastly, it is hard to draw conclusions regarding impact due to a lack of comparisons between the effectiveness of green nudges and the effectiveness of traditional environmental policy tools.

Evaluating the ethical criteria in our framework required a heavy reliance on existing literature from an array of academic fields, e.g., among others, philosophy, law and political science. In many cases, we were unable to evaluate each criterion sufficiently enough to draw concrete conclusions. Even if autonomy, manipulation and transparency were straightforward, the conclusions drawn are nonetheless somewhat subjective. However, subjectivity is unavoidable in a normative evaluation of green nudging and is not always undesirable. For instance, how one weighs the various sub-criteria in terms of importance provides flexibility for policymakers to tailor the framework to specific circumstances and contexts. In our evaluation, we implicitly gave all sub-criteria equal weighting, but one could just as easily emphasise, e.g., the sub-criterion transparency to ensure the visibility of a nudge in society.

The conclusions that can be drawn regarding the autonomy, manipulation and transparency of green default and green social norm nudges reflect the current state of debate. Evaluating the principle of proportionality, on the other hand, required us to draw pre-emptive conclusions. This evaluation, combining ethical concerns with the principles of proportionality, is the first of its kind and provides policymakers with a unique framework encompassing both crucial ethical considerations whilst also considering the legal ramifications of a new policy instrument. Furthermore, the principle of proportionality assessment highlights the aspects of green nudges that are superior to alternative policy measures and weighs these against

¹⁹ More recently, Allcott and Rogers (2014) have shown evidence for a durable effect of social norm nudges for energy conservation.

other ethical implications. The resulting insights help inform policymakers where and in which contexts green nudges may be socially acceptable.

In Section 4.2.3, we mentioned that nudging has been heavily criticised on the basis that it can compromise citizens' ability to form preferences. Bovens (2009) suggests that the coherence (or lack thereof) of the preference structure an individual ends up with after being nudged indicates her autonomy losses. As he puts it, with fragmented preferences, an agent risks being eventually unable to recognize herself in her own actions (pp. 212-214). For example, we might eat sustainably when selecting a meal from a cafeteria that choice architects have designed to steer our behaviour only to revert to bad eating habits when selecting food and snacks in other choice architectures. The issue of fragmentation concerns the development of moral character over time (Selinger & Whyte, 2011, p. 929). The critique of nudges here refers to the notion that if one is not challenged to learn to make good choices in most contexts, one will expect other members of society to take responsibility for their decisions and nudge them away from anything that is bad for them. This could lead to morally lazy, fragmented selves, who are quick to rely on others to ensure their own welfare (Selinger & Whyte, 2011, p. 929).

The concept of the 'fragmented self' is problematic for the legitimacy of nudging because people who do not value personal responsibility may accept forms of government that are not accountable to citizens and in which most decisions about how to modify citizens' behaviour are made behind closed doors without any public input, consent or engagement (Selinger & Whyte, 2011, p. 929). It is possible to imagine an ethical objection that would contend that some nudges do not allow people to build up their own capacities and might even undermine their incentives to do so (Bovens, 2009, pp. 212-214). The more we become familiarized with being nudged, the less we may be bothered by the incremental introduction of more controlling tactics.

In addition to compromising citizens' ability to form preferences, nudging has been heavily criticised for being unfair. Sunstein and Reisch (2013) claim that nudging often works best on the uninformed and uneducated members of society (pp. 400-401). Therefore, a central concern should be the risk that green nudge policies will allow a minority of well-informed citizens to free-ride on the efforts of an less-informed majority (Lehner, Mont & Heiskanen, 2016, p. 175). This is particularly relevant when discussing nudges addressing common goods, such as climate protection. An example of this, for instance, is the green default nudge addressing electricity consumption. One may knowingly opt out of green energy to avoid being nudged, thus enjoying the climatic mitigation benefits provided by the majority using sustainable energy whilst not participating financially. As Goodwin (2012) and Lehner et al. (2016) point out, it is democratically worrying to use nudging to influence the behaviour of those not able to identify a nudge, while allowing those that are able to identify it (and thus avoid it) escape the costs while benefiting from the gains.

Finally, the fairness of green nudging is frequently discussed regarding the distributive impacts of a given nudge.²⁰ Shah, Mullainathan and Shafir (2012) reason that the poor face a relatively higher cognitive load than the rich: struggling with making ends meet on a regular basis consumes scarce cognitive resources (pp. 683-684). Therefore, on the one hand, the poor may benefit by being relieved from cognitive stress when it comes to passive nudges but may lose out by not being able to enjoy the benefits that activating nudges can provide. Additionally, lower income communities with a higher cognitive load may also be more susceptible to green nudges. Analogous reasoning applies to the potential redistributive impacts of different kinds of nudges. For instance, nudges aimed at System 1 may have a larger impact on intuitive thinking consumers, as compared to their more analytically thinking peers (Hagman, Andersson, Västfjäll & Tinghög, 2015, p. 443). Additionally, fairness is relevant as a matter of perception as it is frequently argued that policy measures in general do not work unless they are perceived as fair (Gowdy, 2008, p. 642). This last point illustrates how ethical concerns are inevitably linked to the other two evaluative criteria in our framework. Green nudges deemed dubious by a population may be actively avoided and thus less effective in promoting pro-environmental behaviour.

7 Conclusion

Should we nudge for the environment? The question posed at the onset of this report cannot be answered with a definite yes or no recommendation, and instead deserves a far more nuanced treatment. In this paper, we not only provided an overview and typology of green nudging as a new policy instrument but went a step further by developing an evaluative framework specifically designed for green nudges. Putting this framework to work, we conducted a preliminary assessment of two common nudges: green defaults and green social norms.

Policymakers have different tools at their disposal – economic, regulatory and informational – to entice, coerce and promote environmentally friendly behaviour. However, the historical neglect of behavioural economics, nudge theory and bounded rationality in environmental policymaking has led to suboptimal results in many cases. Overwhelming evidence from the behavioural sciences suggests that ‘Humans’ by nature fall short of the rational choice, *homo economicus* ideal. The logical conclusion is therefore: Environmental policy must be tailored to account for limitations in human cognition and the impact of choice context. Green nudging represents a step in this direction. The results of our preliminary assessment of the efficacy and effectiveness of green nudging can be summarised in the first key conclusion of this report.

²⁰ i.e., the extent to which green nudges redistribute either well-being or freedom among heterogeneous populations.

(1) Green nudging is an effective tool for promoting pro-environmental behaviour both in and outside of a laboratory setting.

For a policy field like environmental policy, where stakes are high and challenges such as climate change immense, nudging is a promising new addition to the toolbox. It represents a cheap and effective means of influencing citizens to act and consume more sustainably, conserve energy, use public transport and waste less food. It is backed up by extensive experimental laboratory research and has been tested successfully in the field. (A *non-exhaustive* summary of empirical evidence for the effectiveness of two types of green nudges can be found in the Annex to this report.) Ends aimed at the common good and environmental integrity may justify the means, especially in societies where there is broad support for ambitious environmental policies. Questions pertaining to how we can reduce our carbon footprint and live more ecologically friendly bother citizens and policymakers alike, and nudge theory offers empirical explanations as to why we have such difficulties changing our behaviours and lifestyles. Crucially, green nudging also represents a multidisciplinary approach to environmental policymaking. Insights from the behavioural sciences and psychology as well as experimentation using RCTs will likely play an increasingly significant role in the future of policymaking.

Nevertheless, contrary to green nudges, conventional policy instruments are not confronted with the type of ethical and psychological concerns that are typically associated with a critical examination of nudging. The traditional contents of the environmental policy toolbox are largely evaluated for their success in protecting environmental goods at the lowest (economic) cost to society, and apart from distributional concerns, ethical considerations tend to fall by the wayside. This brings us to the second key conclusion of this report.

(2) A narrow policy evaluation that omits ethical considerations is insufficient for assessing the potential and suitability of green nudging.

Are nudges manipulative? Do they really retain freedom of choice? Is it fair to make use of people's weaknesses in will and decision making? Should a democratic state be allowed to source out policy architecture to a non-elected 'nudge unit' staff? Is there a danger of citizens losing their ability to make personal decisions because the state decides what is best for them and nudges them in a certain direction? These are among the many questions that must be critically addressed before a public authority pursues nudging in earnest. Moreover, a critical evaluation of ethical aspects *prior* to the implementation of each new green nudge is crucial. As we have observed, much of the existing literature on green nudging is positively biased towards the *one-size-fits-all* use of nudging for environmental aims. We recommend that environmental policymakers perform a complete assessment for each proposed nudge prior to its application, using the criteria of efficacy, effectiveness and ethics as a minimum standard framework for evaluation. This may imply higher costs related to an otherwise cost-effective policy tool, but we contend that it is well-invested public money. It is in the public's interest to comprehensively evaluate an instrument that may have negative ethical and

political consequences, not least of which being a decrease in trust in government institutions. Especially in times of increasing public scepticism about decision making by politicians, it is important to have a transparent and ethically focused approach to behavioural interventions. Otherwise, citizens might, understandably, not see the 'liberal' aspect in 'libertarian paternalism' and fail to recognise the potentially beneficial impact of a nudge-based environmental policy. Finally, it is advisable to review implemented nudges at regular time intervals to make sure that no side effects have occurred in the meantime. This would also help maintain awareness of a nudge's existence to avoid the problem of 'bequeathing' nudges to new generations of policymakers or administrative staff.

A critical and permanent debate on green nudges is a required step towards finding solutions to the ethical and practical concerns outlined in this paper. Hopefully, the evaluative framework developed here can be used by policymakers and practitioners to produce effective behavioural interventions that treat citizens with respect by upholding their dignity and considering their interests. This is in no way different than how Thaler and Sunstein first envisioned the potential of nudges to do good. It is paramount that an on-going discourse continues to scrutinise nudges from an ethical standpoint, especially considering the positive reception the policy tool has received by governments worldwide and the speed at which it is being implemented.

8 Annex: Overview of green nudges

Table 5: Empirical examples of green default nudges

Reference	Behaviour	Details of study	Effect
Araña and León (2013)	Carbon offsetting with purchase	Large scale field experiment (1600 participants)	Default (opt-out) condition led to an average 22.7% increase in willingness to pay for carbon offset (25.5€ vs. 31.29€)
Brown, Johnstone, Haščič, Vong and Barascud (2013)	Conserving energy on heating and cooling in the US	Medium scale field experiment (87 offices)	On average a 1.0°C decrease in the default led to a reduction of 0.38°C in the chosen thermostat setting
Dinner et al. (2011)	Buying energy efficient light bulbs in the US	Medium scale online choice experiment and survey (120-200 participants)	Default led to a 23.6%, 22.9% and 39% increases in choice of more efficient light bulbs across three experiments
Egebark and Ekström (2016)	Conserving paper at a university in Sweden	Medium scale natural field experiment (18 departments at a major Swedish university)	Default double-sided printing led to 15% decrease in paper consumption, effect remains six months following intervention
Hedlin and Sunstein (2016)	Choosing renewable electricity provider in US	Large scale online choice experiment and survey (1200 participants)	Green energy default led to a 13% increase in renewable energy choice; active choice between grey and green energy led to 20% increase
Kallbekken and Saelen (2013)	Reducing food waste at hotels in Norway	Large scale field experiment (52 hotels)	Default plate size led to 19.5% decrease in food waste
Löfgren, Martinsson, Hennlock and Sterner (2012)	Carbon offsetting with purchase	Medium scale field experiment (240 participants)	No significant impact of default or active choice on decision to offset carbon emissions
Momsen and Stoerk (2014)	Choosing renewable electricity provider in Germany	Medium scale choice experiment, online survey (475 participants)	44.6% increase due to default

Reference	Behaviour	Details of study	Effect
Pichert and Katsikopoulos (2007)	Choosing renewable electricity provider in Germany	Mixed methods: two natural experiments (2500 participants, 1500000 customers); a laboratory choice experiments (225 participants)	Natural experiment #1: 1669 out of 1683 meters (99%) stay with (default) green provider after given the option to change Natural experiment #2: 94% stay with default tariff Laboratory experiment #1: 17% increase due to default, effect size = 0.26 (medium)
Toft, Schuitema and Thøgersen (2014), described in Ölander and Thøgersen (2014)	Choosing a home smart meter in Denmark, Norway and Switzerland	Medium scale choice experiment (1800 participants) and field experiment (140 households)	Choice experiment: 18% increase in smart meter uptake in default condition, default condition increases the chance of smart meter uptake by 2.65 Field experiment: default (opt-out) condition led to 46% increase in smart meter uptake
Vetter and Kutzner (2016)	Choosing renewable electricity provider in Germany	Medium scale choice experiment (600 participants)	Participants four times more likely to choose green energy under green default

Table 6: Empirical examples of green social norm nudges

Reference	Behaviour	Details of study	Effect
Allcott (2011)	Residential electricity consumption in US	Large scale field experiment (600000 households); two-year intervention period	Average 2% reduction in consumption; 0.62kWh/day saved
Allcott and Rogers (2014)	Residential electricity consumption in US	Large scale field experiment (230000 households); four-year intervention period, longevity of nudge was studied after half of the treatment group stopped the intervention	Average 1-1.3% immediate reduction in consumption; 0.3-0.4 kWh/day saved, after two months this drops to 0.2 kWh/day; discontinuing nudge leads to a decay of effect by 10-20% per year

Reference	Behaviour	Details of study	Effect
Ayres, Raseman and Shih (2013)	Residential electricity consumption in US	Large scale natural field experiment (170000 households)	Average 1.2-2.1% reduction in consumption
Brandon and Lewis (1999)	Gas and electricity consumption in UK households	Large scale field experiment (1000 households)	4.6% decrease in energy usage with comparative norm intervention
Carrus, Bonnes, Fornara, Passafaro and Tronu (2009)	Residential recycling behaviour in Italy	Medium scale survey (300 participants)	Descriptive local social norms correlated with behavioural intention ($r = 0.49$)
Cialdini, Reno and Kallgren (1990)	Littering in public places in the US	Medium scale field experiment (150-400 individuals)	Multiple experiments: range from 7-40% decrease in littering behaviour
Cialdini, Demaine, Sagarin, Barrett, Rhoads and Winter (2006)	Theft of petrified wood from a US National Park	Large scale field experiment (2600 individuals)	No control group. Study looked at interaction between type of norm (i.e. injunctive/descriptive) and strength (negatively/positively worded). A strong injunctive norm had the greatest effect: 1.67% theft (proportion of wood pieces stolen). Strong descriptive norm had least effect: 7.92%
Costa and Kahn (2010)	Residential electricity consumption in US	Large scale field experiment (85000 individuals)	Average 2% reduction in consumption (1.7% for conservatives; 2.4% for liberals)
Dolan and Metcalfe (2015)	Residential electricity consumption in UK	Medium scale field experiment (569 households)	Average 6% reduction in consumption
Fornara, Carrus, Passafaro and Bonnes (2011)	Residential recycling behaviour in Italy	Medium scale survey (452 participants)	Descriptive local norms predicted behavioural intention mediated by perceived behavioural control ($R^2 = 0.40$); injunctive norms account for less variance in model
Goldstein, Cialdini and Griskevicius, (2008)	Reusing towels in US hotels	Medium scale field experiment (1058 instances of potential towel reuse in 190 rooms)	Specific room social norm led to a 9% increase in towel reuse compared to a standard environmental message

Reference	Behaviour	Details of study	Effect
Graffeo, Ritov, Bonini and Hadjichristidis (2015)	Intended household energy conservation in Israel	Medium scale choice experiment (300 individuals)	Social norm interventions led to an intention to conserve energy by 12.1-34.4% compared to a control group (depending on manipulation)
Kallbekken and Saelen (2013)	Reducing food waste in hotel restaurants in Norway	Large scale field experiment (52 hotels)	Social cue intervention led to a 20.5% decrease in waste
Kormos, Gifford and Brown (2015)	Self-reported private vehicle use	Medium scale field experiment (78 participants; month-long intervention)	Non-significant main effect of social norm on overall sustainable transport behaviour; positive linear relation between sustainable transportation and strength of descriptive norm; marginally-significant main effect of norm on commuting behaviour
Kuhfuss, Préget, Thoyer and Hanley (2016)	Signing of agri-environmental contracts to limit pesticide use in France	Medium scale choice experiment (317 farmers)	Collective bonus (implicit social norm) leads to farmers' willingness to accept a bonus of 30€/ha/year less than control group, a 22% difference
McCalley and Midden (2002)	Energy conservation behaviour	Medium scale laboratory experiment (100 participants)	Self-set group goal led to 21% reduction in energy consumption
Midden, Meter, Weenig and Zieverink (1983)	Residential gas and electricity consumption Netherlands	Medium scale field experiment (91 households)	Non-significant effect of comparative norm; comparative norm plus incentive led to reductions of 19.4% for electricity and 17.5% for gas
Nigbur, Lyons and Uzzell (2010)	Residential recycling behaviour in UK	Medium scale survey and natural experiment (527 participants)	Descriptive social norm significant predictor of intended and actual behaviour
Nolan, Schultz, Cialdini, Goldstein and Griskevicius (2008)	Residential electricity consumption in US	Medium scale field experiment (290 households)	Social norm messaging led to 10% reduction in consumption
Schultz (1999)	Residential recycling behaviour in US	Medium scale field experiment (120 households)	Social norm messaging led to 19% increase in recycling behaviour

Reference	Behaviour	Details of study	Effect
Schultz, Nolan, Ci-aldini, Goldstein and Griskevicius (2007)	Residential electricity consumption in US	Medium scale field experiment (270 households)	Descriptive norm led to decrease of 1.22kWh/day for more-than-average energy users; for less-than-average users it led to a 0.89kWh/day increase; boomerang effect disappears with inclusion of injunctive norm: 2.5% decrease for less-than-average users and 8.3% for more-than-average
Schultz, Khazian and Zaleski (2008)	Reusing towels in US hotels	Large scale field experiment (60 hotels; 800-2300 guests per experiment)	Average towels replaced per stay Experiment #2 Combined injunctive and descriptive norm: 1.74 Control group: 2.32 Experiment #3 Specific norm: 2.19 Generic norm: 2.02 Control group: 2.44

9 References

- Alemanno, A., Sibony, A.-L., & Sunstein, C. R. (Eds.). (2015). *Nudge and the law: a European perspective*. Oxford, United Kingdom: Hart Publishing.
- Alexy, R. (2003). Constitutional Rights, Balancing, and Rationality. *Ratio Juris*, 16(2), 131-140. <https://doi.org/10.1111/1467-9337.00228>
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9), 1082-1095.
- Allcott, H., & Mullainathan, S. (2010). Behavior and Energy Policy. *Science*, 327(5970), 1204-1205. <https://doi.org/10.1126/science.1180775>
- Allcott, H., & Rogers, T. (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, 104(10), 3003-3037. <https://doi.org/10.1257/aer.104.10.3003>
- Amir, O., & Lobel, O. (2008). Stumble, predict, nudge: how behavioral economics informs law and policy. *Columbia Law Review*, 108(8), 2098-2138.
- Araña, J. E., & León, C. J. (2013). Can Defaults Save the Climate? Evidence from a Field Experiment on Carbon Offsetting Programs. *Environmental and Resource Economics*, 54(4), 613-626. <https://doi.org/10.1007/s10640-012-9615-x>
- Ayres, I., Raseman, S., & Shih, A. (2013). Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. *Journal of Law, Economics, and Organization*, 29(5), 992-1022. <https://doi.org/10.1093/jleo/ews020>
- Bamberg, S. (2002). Effects of implementation intentions on the actual performance of new environmentally friendly behaviours - Results of two field experiments. *Journal of Environmental Psychology*, 22(4), 399-411. <https://doi.org/10.1006/jevp.2002.0278>
- Barton, A., & Grüne-Yanoff, T. (2015). From Libertarian Paternalism to Nudging—and Beyond. *Review of Philosophy and Psychology*, 6(3), 341-359. <https://doi.org/10.1007/s13164-015-0268-x>
- Becker, L. J. (1978). Joint effect of feedback and goal setting on performance: A field study of residential energy conservation. *Journal of Applied Psychology*, 63(4), 428-433. <https://doi.org/10.1037/0021-9010.63.4.428>
- Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). How are preferences revealed? *Journal of Public Economics*, 92(8-9), 1787-1794. <https://doi.org/10.1016/j.jpubeco.2008.04.010>
- Binder, M. (2014). A constitutional paradigm is not enough—would sovereign citizens really agree to manipulative nudges?—A reply to Christian Schubert. *Journal of Evolutionary Economics*, 24(5), 1115-1120. <https://doi.org/10.1007/s00191-014-0377-1>

- BMUB. (2017). *Umweltbewusstsein in Deutschland 2016: Ergebnisse einer repräsentativen Bevölkerungsumfrage* (No. 3715 16 103 0). Berlin: Bundesministeriums für Umwelt, Naturschutz, Bau und Reaktorsicherheit (BMUB).
- Borrás, S., & Edquist, C. (2013). The choice of innovation policy instruments. *Technological Forecasting and Social Change*, 80(8), 1513-1522. <https://doi.org/10.1016/j.techfore.2013.03.002>
- Bovens, L. (2009). The Ethics of Nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference Change* (pp. 207-219). Dordrecht: Springer Netherlands. Retrieved from http://www.springerlink.com/index/10.1007/978-90-481-2593-7_10
- Brandon, G., & Lewis, A. (1999). Reducing household energy consumption: A qualitative and quantitative field study. *Journal of Environmental Psychology*, 19, 75-85.
- Brown, O. (2007). *Fighting climate change: Human solidarity in a divided world* (Background paper). New York, NY: United Nations Development Programme.
- Brown, Z., Johnstone, N., Haščič, I., Vong, L., & Barascud, F. (2013). Testing the effect of defaults on the thermostat settings of OECD employees. *Energy Economics*, 39, 128-134. <https://doi.org/10.1016/j.eneco.2013.04.011>
- Carroll, J., Lyons, S., & Denny, E. (2014). Reducing household electricity demand through smart metering: The role of improved information about energy saving. *Energy Economics*, 45, 234-243. <https://doi.org/10.1016/j.eneco.2014.07.007>
- Carrus, G., Bonnes, M., Fornara, F., Passafaro, P., & Tronu, G. (2009). Planned behavior and “local” norms: an analysis of the space-based aspects of normative ecological behavior. *Cognitive Processing*, 10(S2), 198-200. <https://doi.org/10.1007/s10339-009-0292-9>
- Centre d’analyse stratégique. (2011). *Green nudges: new incentives for ecological behavior* (No. 216). Paris: Centre d’analyse stratégique.
- Christman, J. (2015). Autonomy in Moral and Political Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition). Retrieved from <http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/>
- Cialdini, R. B. (2005). Don’t throw in the towel: Use social influence research. *American Psychological Society Observer*, 18, 33-34.
- Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. New York, NY US: Harper Business.
- Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social Influence*, 1(1), 3-15. <https://doi.org/10.1080/15534510500181459>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, 55(1), 591-621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>

- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. *Journal of Personality and Social Psychology*, 58(6), 1015-1026.
- Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, 16(20), 2349-2380. [https://doi.org/10.1002/\(SICI\)1097-0258\(19971030\)16:20<2349::AID-SIM667>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-0258(19971030)16:20<2349::AID-SIM667>3.0.CO;2-E)
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9-13. <https://doi.org/10.3102/0013189X032001009>
- Conly, S. (2013). *Against autonomy: justifying coercive paternalism*. Cambridge: Cambridge University Press.
- Costa, D. L., & Kahn, M. E. (2013). Energy Conservation “Nudges” and Environmentalist Ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3), 680-702. <https://doi.org/10.1111/jeea.12011>
- Crosan, R., & Treich, N. (2014). Behavioral Environmental Economics: Promises and Challenges. *Environmental and Resource Economics*, 58(3), 335-351. <https://doi.org/10.1007/s10640-014-9783-y>
- Dams, J., Ettl, A., Greive, M., & Zschäpitz, H. (2015, March 12). Merkel will die Deutschen durch Nudging erziehen. *Die Welt*. Retrieved from <http://www.welt.de/138326984>
- Dietz, T., Ostrom, E., & Stern, P. C. (2003). The struggle to govern the commons. *Science*, 302(5652), 1907-1912.
- Dinner, I., Johnson, E. J., Goldstein, D. G., & Liu, K. (2011). Partitioning default effects: Why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17(4), 332-341. <https://doi.org/10.1037/a0024354>
- Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). *MINDSPACE: Influencing behaviour through public policy*. London, UK: Institute for Government.
- Dolan, P., & Metcalfe, R. D. (2015). Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2589269>
- Dworkin, G. (1988). *The theory and practice of autonomy*. Cambridge ; New York: Cambridge University Press.
- Egebark, J., & Ekström, M. (2016). Can indifference make the world greener? *Journal of Environmental Economics and Management*, 76, 1-13. <https://doi.org/10.1016/j.jeem.2015.11.004>

- European Commission. (2012). *Directive 2012/27/EU of the European Parliament and of the Council on energy efficiency* (Official Journal of the European Union No. OJ L 315).
- European Commission. (2016). *Behavioural Insights Applied to Policy: European Report 2016* (No. EUR 277726 EN). European Commission.
- European Council. (2002). *Decision of the European Parliament and of the Council of 22 July 2002 laying down the sixth community environment action programme* (No. 1600/2002/EC). Brussels.
- Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, 59(1), 255-278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223-241. <https://doi.org/10.1177/1745691612460685>
- Félonneau, M.-L., & Becker, M. (2008). Pro-environmental attitudes and behavior: Revealing perceived social desirability. *RIPSO Revue Internationale de Psychologie Sociale*, 21(4), 25-53.
- Fischer, M., & Lotz, S. (2014). *Is Soft Paternalism Ethically Legitimate? - The Relevance of Psychological Processes for the Assessment of Nudge-Based Policies* (Working Paper No. 5-2). Cologne, Germany: Cologne Graduate School in Management, Economics and Social Sciences.
- Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventative Medicine*, 15(5), 451-474.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... Ji, P. (2005). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science*, 6(3), 151-175. <https://doi.org/10.1007/s11121-005-5553-y>
- Fornara, F., Carrus, G., Passafaro, P., & Bonnes, M. (2011). Distinguishing the sources of normative influence on proenvironmental behaviors: The role of local norms in household waste recycling. *Group Processes & Intergroup Relations*, 14(5), 623-635. <https://doi.org/10.1177/1368430211408149>
- Gerber, A. S., & Green, D. P. (2011). *Field experiments and natural experiments*. Oxford, UK: Oxford University Press.
- Geva-May, I., & Pal, L. A. (1999). Good fences make good neighbors: Policy evaluation and policy analysis - exploring the differences. *Evaluation*, 5(3), 259-277.
- Gigerenzer, G. (2015). On the Supposed Evidence for Libertarian Paternalism. *Review of Philosophy and Psychology*, 6(3), 361-383. <https://doi.org/10.1007/s13164-015-0248-1>

- Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why Don't We See More Translation of Health Promotion Research to Practice? Rethinking the Efficacy-to-Effectiveness Transition. *American Journal of Public Health, 93*(8), 1261-1267. <https://doi.org/10.2105/AJPH.93.8.1261>
- Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research, 35*(3), 472-482. <https://doi.org/10.1086/586910>
- Goodwin, T. (2012). Why We Should Reject "Nudge." *Politics, 32*(2), 85-92. <https://doi.org/10.1111/j.1467-9256.2012.01430.x>
- Gowdy, J. M. (2008). Behavioral economics and climate change policy. *Journal of Economic Behavior & Organization, 68*(3-4), 632-644. <https://doi.org/10.1016/j.jebo.2008.06.011>
- Graffeo, M., Ritov, I., Bonini, N., & Hadjichristidis, C. (2015). To make people save energy tell them what others do but also who they are: a preliminary study. *Frontiers in Psychology, 6*. <https://doi.org/10.3389/fpsyg.2015.01287>
- Grasmick, H. G., Bursik, R. J., & Kinsey, K. A. (1991). Shame and Embarrassment as Deterrents to Noncompliance with the Law: The Case of an Antilittering Campaign. *Environment and Behavior, 23*(2), 233-251. <https://doi.org/10.1177/0013916591232006>
- Gribbons, B., & Herman, J. (1997). *True and Quasi-Experimental Designs*. Washington, DC: ERIC/AE Digest.
- Grossman, S. J., & Hart, O. D. (1980). Takeover bids, the free-rider problem, and the theory of the corporation. *Bell Journal of Economics, 11*(1), 42-64.
- Grüne-Yanoff, T. (2007). Bounded Rationality. *Philosophy Compass, 2*(3), 534-563. <https://doi.org/10.1111/j.1747-9991.2007.00074.x>
- Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge Versus Boost: How Coherent are Policy and Theory? *Minds and Machines, 26*(1-2), 149-183. <https://doi.org/10.1007/s11023-015-9367-9>
- Hacker, P. (2016). Nudging and Autonomy: A Philosophical and Legal Appraisal (Forthcoming). *Handbook of Research Methods in Consumer Law*.
- Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public Views on Policies Involving Nudges. *Review of Philosophy and Psychology, 6*(3), 439-453. <https://doi.org/10.1007/s13164-015-0263-2>
- Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove. *European Journal of Risk Regulation, 1*, 1-20.
- Hansen, P. G., & Jespersen, A. (2013). Nudge and the Manipulation of Choice. A Framework for the Responsible Use of Nudge Approach to Behaviour Change in Public Policy. *European Journal of Risk Regulation, 1*, 3-28.

- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243-1248. <https://doi.org/10.1126/science.162.3859.1243>
- Hausman, D. M., & Welch, B. (2010). Debate: To Nudge or Not to Nudge*. *Journal of Political Philosophy*, 18(1), 123-136. <https://doi.org/10.1111/j.1467-9760.2009.00351.x>
- Hedlin, S., & Sunstein, C. R. (2016). Does Active Choosing Promote Green Energy Use? Experimental Evidence. *Ecology Law Quarterly*, 43(107-141), 1-37.
- Hildén, M., Lepola, J., Mickwitz, P., Mulders, A., Palosaari, M., Similä, J., ... Vedung, E. (2002). *Evaluation of environmental policy instruments: A case study of the Finnish pulp and paper and chemical industries*. Finnish Environment Institute.
- House, J., & Lyons, E. (2013). *Towards a Taxonomy of Nudging Strategies* (Behavioural Economics in Action). Toronto: Rotman School of Management University of Toronto.
- Jann, W., & Wegrich, K. (2009). Phasenmodelle und Politikprozesse: der policy cycle. *Lehrbuch Der Politikfeldanalyse*, 2(2), 75-114.
- Joachain, H., & Klopfert, F. (2014). Smarter than metering? Coupling smart meters and complementary currencies to reinforce the motivation of households for energy savings. *Ecological Economics*, 105, 89-96. <https://doi.org/10.1016/j.ecolecon.2014.05.017>
- John, P., Cotterill, S., Moseley, A., Richardson, L., Smith, G., Stoker, G., & Wales, C. (2013). *Nudge, nudge, think, think: experimenting with ways to change civic behaviour* (Revised paperback edition). London New Delhi New York Sydney: Bloomsbury.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697-720. <https://doi.org/10.1037/0003-066X.58.9.697>
- Kahneman, D. (2012). *Thinking, fast and slow*. London: Penguin Books.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341-350.
- Kallbekken, S., & Sælen, H. (2013). "Nudging" hotel guests to reduce food waste as a win-win environmental measure. *Economics Letters*, 119(3), 325-327. <https://doi.org/10.1016/j.econlet.2013.03.019>
- Kallbekken, S., Sælen, H., & Hermansen, E. A. T. (2013). Bridging the Energy Efficiency Gap: A Field Experiment on Lifetime Energy Costs and Household Appliances. *Journal of Consumer Policy*, 36(1), 1-16. <https://doi.org/10.1007/s10603-012-9211-z>
- Kinzig, A. P., Ehrlich, P. R., Alston, L. J., Arrow, K., Barrett, S., Buchman, T. G., ... Saari, D. (2013). Social Norms and Global Environmental Challenges: The Complex Interaction of Behaviors, Values, and Policy. *BioScience*, 63(3), 164-175. <https://doi.org/10.1525/bio.2013.63.3.5>

- Kollmuss, A., & Agyeman, J. (2002). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, 8(3), 239-260. <https://doi.org/10.1080/13504620220145401>
- Kormos, C., Gifford, R., & Brown, E. (2015). The Influence of Descriptive Social Norm Information on Sustainable Transportation Behavior: A Field Experiment. *Environment and Behavior*, 47(5), 479-501. <https://doi.org/10.1177/0013916513520416>
- Kuhfuss, L., Préget, R., Thoyer, S., & Hanley, N. (2016). Nudging farmers to enrol land into agri-environmental schemes: the role of a collective bonus. *European Review of Agricultural Economics*, 43(4), 609-636. <https://doi.org/10.1093/erae/jbv031>
- Lee, L., Amir, O., & Ariely, D. (2009). In Search of Homo Economicus: Cognitive Noise and the Role of Emotion in Preference Consistency. *Journal of Consumer Research*, 36(2), 173-187. <https://doi.org/10.1086/597160>
- Lehner, M., Mont, O., & Heiskanen, E. (2016). Nudging - A promising tool for sustainable consumption behaviour? *Journal of Cleaner Production*, 134, 166-177. <https://doi.org/10.1016/j.jclepro.2015.11.086>
- Lepenes, R., & Matecka, M. (2015). The Institutional Consequences of Nudging - Nudges, Politics, and the Law. *Review of Philosophy and Psychology*, 6(3), 427-437. <https://doi.org/10.1007/s13164-015-0243-6>
- Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: constraining our future selves to ameliorate global climate change. *Policy Sciences*, 45(2), 123-152. <https://doi.org/10.1007/s11077-012-9151-0>
- Lewis, M. A., & Neighbors, C. (2006). Social Norms Approaches Using Descriptive Drinking Norms Education: A Review of the Research on Personalized Normative Feedback. *Journal of American College Health*, 54(4), 213-218. <https://doi.org/10.3200/JACH.54.4.213-218>
- Linder, S. H., & Peters, B. G. (1998). The study of policy instruments: four schools of thought. In B. G. Peters & F. K. M. van Nispen (Eds.), *Public policy instruments: evaluating the tools of public administration* (pp. 33-45). Cheltenham, UK ; Northampton, MA: Edward Elgar Publishing.
- Löfgren, Å., Martinsson, P., Hennlock, M., & Sterner, T. (2012). Are experienced people affected by a pre-set default option—Results from a field experiment. *Journal of Environmental Economics and Management*, 63(1), 66-72. <https://doi.org/10.1016/j.jeem.2011.06.002>
- Lundqvist, L. J. (1996). Environmental Politics in the Nordic Countries: Policy, Organisation, and Capacity. In P. M. Christiansen (Ed.), *Governing the Environment: Politics, Policy, and Organization in the Nordic Countries* (pp. 13-27). Copenhagen: Nordic Council of Ministers.

- McCalley, L. ., & Midden, C. J. . (2002). Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of Economic Psychology*, 23(5), 589-603. [https://doi.org/10.1016/S0167-4870\(02\)00119-8](https://doi.org/10.1016/S0167-4870(02)00119-8)
- Michalek, G., Meran, G., Schwarze, R., & Yildiz, Ö. (2015). *Nudging as a new “soft” tool in environmental policy: An analysis based on insight from cognitive and social psychology* (Discussion Paper No. 21). Frankfurt (Oder): Europa Universität.
- Mickwitz, P. (2003). A framework for evaluating environmental policy instruments context and key concepts. *Evaluation*, 9(4), 415-436.
- Midden, C. J. H., Meter, J. E., Weenig, M. H., & Zieverink, H. J. . (1983). Using feedback, reinforcement and information to reduce energy consumption in households: A field-experiment. *Journal of Economic Psychology*, 3(1), 65-86.
- Mill, J. S. (1863). *On Liberty*. Boston: Ticknor and Fields.
- Mitchell, G. (2005). Libertarian Paternalism Is an Oxymoron. *Northwestern University Law Review*, 99(3). Retrieved from <http://papers.ssrn.com/abstract=615562>
- Mols, F., Haslam, S. A., Jetten, J., & Steffens, N. K. (2015). Why a nudge is not enough: A social identity critique of governance by stealth: Why a nudge is not enough. *European Journal of Political Research*, 54(1), 81-98. <https://doi.org/10.1111/1475-6765.12073>
- Momsen, K., & Stoerk, T. (2014). From intention to action: Can nudges help consumers to choose renewable energy? *Energy Policy*, 74, 376-382. <https://doi.org/10.1016/j.enpol.2014.07.008>
- Mont, O., Lehner, M., Heiskanen, E., Sverige, & Naturvårdsverket. (2014). *Nudging: a tool for sustainable behaviour?* Stockholm: Swedish Environmental Protection Agency.
- Moseley, A., & Stoker, G. (2013). Nudging citizens? Prospects and pitfalls confronting a new heuristic. *Resources, Conservation and Recycling*, 79, 4-10. <https://doi.org/10.1016/j.resconrec.2013.04.008>
- Nagatsu, M. (2015). Social Nudges: Their Mechanisms and Justification. *Review of Philosophy and Psychology*, 6(3), 481-494. <https://doi.org/10.1007/s13164-015-0245-4>
- Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin*, 126(6), 964.
- Nigbur, D., Lyons, E., & Uzzell, D. (2010). Attitudes, norms, identity and environmental behaviour: Using an expanded theory of planned behaviour to predict participation in a kerbside recycling programme. *British Journal of Social Psychology*, 49(2), 259-284. <https://doi.org/10.1348/014466609X449395>
- Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative Social Influence is Underdetected. *Personality and Social Psychology Bulletin*, 34(7), 913-923. <https://doi.org/10.1177/0146167208316691>

- Nys, T. R., & Engelen, B. (2016). Judging Nudging: Answering the Manipulation Objection. *Political Studies*. <https://doi.org/10.1177/0032321716629487>
- Office of the Press Secretary. (2015). *Press release: President Obama Signs Executive Order; White House Announces New Steps to Improve Federal Programs by Leveraging Research Insights*. Washington, DC: The White House.
- Ölander, F., & Thøgersen, J. (2014). Informing Versus Nudging in Environmental Policy. *Journal of Consumer Policy*, 37(3), 341-356. <https://doi.org/10.1007/s10603-014-9256-2>
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge, England: Cambridge University Press.
- Ostrom, E. (1998). A Behavioral approach to the rational choice theory of collective action. *American Political Science Review*, 92(1), 1-22.
- Ostrom, E. (2003a). Toward a Behavioral Theory Linking Trust, Reciprocity, and Reputation. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: interdisciplinary lessons from experimental research* (pp. 19-79). New York: Russell Sage Foundation.
- Ostrom, E. (2003b). Toward a Behavioral Theory Linking Trust, Reciprocity, and Reputation. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: interdisciplinary lessons from experimental research* (pp. 19-77). New York: Russell Sage Foundation.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 14, 123-192.
- Pichert, D., & Katsikopoulos, K. V. (2008). Green defaults: Information presentation and pro-environmental behaviour. *Journal of Environmental Psychology*, 28(1), 63-73. <https://doi.org/10.1016/j.jenvp.2007.09.004>
- Raihani, N. J. (2013). Nudge politics: efficacy and ethics. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00972>
- Reisch, L., & Hagen, K. (2011). Kann der Konsumwandel gelingen? Chancen und Grenzen einer verhaltensökonomisch basierten sozialen Regulierung. In L. Heidbring, I. Schmidt, & B. Ahaus (Eds.), *Die Verantwortung des Konsumenten: über das Verhältnis von Markt, Moral und Konsum* (pp. 221-243). Frankfurt am Main: Campus Verlag.
- Schnellenbach, J. (2012). Nudges and norms: On the political economy of soft paternalism. *European Journal of Political Economy*, 28(2), 266-277. <https://doi.org/10.1016/j.ejpoleco.2011.12.001>
- Schubert, C. (2015). *On the Ethics of Public Nudging: Autonomy and Agency* (Working Paper). Kassel: University of Kassel. Retrieved from Available at SSRN: <http://ssrn.com/abstract=2672970>
- Schubert, C. (2016). Green Nudges: Do They Work? Are They Ethical? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2729899>

- Schultz, P. W. (1999). Changing behavior with normative feedback interventions: A field experiment on curbside recycling. *Basic and Applied Social Psychology*, 21, 25-38.
- Schultz, W. P., Khazian, A. M., & Zaleski, A. C. (2008). Using normative social influence to promote conservation among hotel guests. *Social Influence*, 3(1), 4-23. <https://doi.org/10.1080/15534510701755614>
- Schultz, W. P., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science*, 18(5), 429-434. <https://doi.org/10.1111/j.1467-9280.2007.01917.x>
- Schweizer, M. (2016). Nudging and the Principle of Proportionality. In K. Mathis & A. Tor (Eds.), *Nudging - Possibilities, Limitations and Applications in European Law and Economics* (pp. 93-119). Cham: Springer International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-319-29562-6_7
- Selinger, E., & Whyte, K. (2011). Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture: Practice and Ethics of Choice Architecture. *Sociology Compass*, 5(10), 923-935. <https://doi.org/10.1111/j.1751-9020.2011.00413.x>
- Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some Consequences of Having Too Little. *Science*, 338(6107), 682-685. <https://doi.org/10.1126/science.1222426>
- Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), 99-118.
- Simon, H. A. (1957). *Models of Man: Social and Rational*. New York: John Wiley and Sons, Inc.
- Simon, H. A. (1972). Theories of bounded rationality. In J. Marschak, C. B. McGuire, R. Radner, & K. J. Arrow (Eds.), *Decision and organization. A volume in honor of Jacob Marschak*. Amsterdam: North-Holland Pub. Co.
- Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397-420). New York: Cambridge University Press.
- Smith, N. C., Goldstein, D. G., & Johnson, E. J. (2013). Choice Without Awareness: Ethical and Policy Implications of Defaults. *Journal of Public Policy & Marketing*, 32(2), 159-172. <https://doi.org/10.1509/jppm.10.114>
- Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology*, 29(3), 309-317. <https://doi.org/10.1016/j.jenvp.2008.10.004>
- Stern, P. C. (2011). Contributions of psychology to limiting climate change. *American Psychologist*, 66(4), 303-314.
- Sunstein, C. R. (2013). Deciding by default. *University of Pennsylvania Law Review*, 162(1), 1-28.

- Sunstein, C. R. (2014). *Why nudge?: the politics of libertarian paternalism*. New Haven: Yale University Press.
- Sunstein, C. R. (2015). The Ethics of Nudging. *Yale Journal on Regulation*, 32(2).
- Tate, K., Stewart, A. J., & Daly, M. (2014). Influencing green behaviour through environmental goal priming: The mediating role of automatic evaluation. *Journal of Environmental Psychology*, 38, 225-232. <https://doi.org/10.1016/j.jenvp.2014.02.004>
- Thaler, R. H. (1980). Towards a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1, 39-60.
- Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3), 199-214.
- Thaler, R. H. (2000). From Homo Economicus to Homo Sapiens. *Journal of Economic Perspectives*, 14(1), 133-141.
- Thaler, R. H. (2016). *Misbehaving: the making of behavioural economics*. United States of America: Norton paperback 2016.
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review: Papers and Proceedings*, (93), 175-179.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.
- Toft, M. B., Schuitema, G., & Thøgersen, J. (2014). The importance of framing for consumer acceptance of the Smart Grid: A comparative study of Denmark, Norway and Switzerland. *Energy Research & Social Science*, 3, 113-123. <https://doi.org/10.1016/j.erss.2014.07.010>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131.
- Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, 106(4), 1039-1061. <https://doi.org/10.2307/2937956>
- UK Science and Technology Committee. (2011). *Science and Technology Select Committee 2nd Report of Session 2010-12 - Behaviour Change* (House of Lords Paper No. 179). London: House of Lords. Retrieved from <http://www.publications.parliament.uk/pa/ld201012/ldselect/ldsctech/179/179.pdf>
- van Aaken, A. (2015). *Constitutional Limits to Nudging: A Proportionality Assessment* (Working Paper No. 2015-3) (pp. 1-28). University of St. Gallen Law & Economics. Retrieved from <http://www.ssrn.com/abstract=2606189>
- van der Heijden, J., & Kusters, M. (2015). From Mechanism to Virtue: Evaluating Nudge-Theory. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2620082>

- Vedung, E. (1998). Policy instruments: typologies and theories. In M.-L. Bemelmans-Videc, R. C. Rist, & E. Vedung (Eds.), *Carrots, sticks & sermons: policy instruments and their evaluation* (pp. 21-58). New Brunswick, N.J., U.S.A: Transaction Publishers.
- Vetter, M., & Kutzner, F. (2016). Nudge me if you can - how defaults and attitude strength interact to change behavior. *Comprehensive Results in Social Psychology*, 1-27. <https://doi.org/10.1080/23743603.2016.1139390>
- Voyer, B. (2015). "Nudging" behaviours in healthcare: Insights from behavioural economics. *British Journal of Healthcare Management*, 21(3), 130-135. <https://doi.org/10.12968/bjhc.2015.21.3.130>
- Wansink, B. (2004). Environmental factors that increase the food uptake and consumption volume of unknowing consumers. *Annual Review of Nutrition*, 24(1), 455-479. <https://doi.org/10.1146/annurev.nutr.24.012003.132140>
- Weber, E. U., & Stern, P. C. (2011). Public understanding of climate change in the United States. *American Psychologist*, 66(4), 315-328. <https://doi.org/10.1037/a0023253>
- Weible, C. M. (2014). Introduction. In P. A. Sabatier & C. M. Weible (Eds.), *Theories of the policy process* (Third edition, pp. 3-24). Boulder, CO: Westview Press, a member of the Persus Books Group.
- White, M. D. (2013). How Behavioral Economics Met Law and Economics and Begat Nudge. In M. D. White, *The Manipulation of Choice* (pp. 43-59). New York: Palgrave Macmillan US. Retrieved from http://link.springer.com/10.1057/9781137313577_3
- Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies*, 61(2), 341-355. <https://doi.org/10.1111/j.1467-9248.2012.00974.x>
- Yeung, K. (2012). Nudge as Fudge: Nudge as Fudge. *The Modern Law Review*, 75(1), 122-148. <https://doi.org/10.1111/j.1468-2230.2012.00893.x>