Ethics of Green Nudging

PhD Thesis

Sukanya Shukla, M.S.

Dublin City University

School of Theology, Philosophy, and Music

Supervised by Dr Fiachra O'Brolchain, Dr Pádraig Murphy, and Prof Patrick Brereton

August, 2026

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original and have conformed to the regulations on the use and declaration of Generative AI, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Sukanya Shukla ID No: 21266953 Date: August 20th, 2026

Reflection on the Use of AI for Grammar-Checking and Citations

Tools	Duration	Prompts	Outputs
Grammarly	02/23 to 09/25	None.	Grammar Checking, Punctuations.
Co-Pilot	06/25 to 09/25	Indentation/formatting related specific questions about reference guidelines, Specific queries about using Word and Google Docs (retaining footnotes during transfer from one doc to another, using citation extensions.)	-

Reflection: Gen AI has been used minimally and responsibly in accordance with the university guidelines. Only Microsoft Co-Pilot was used, as recommended in the guidelines. Neither mine nor anyone else's intellectual property, data, or documents were uploaded to any Gen AI tools. It was used for specific pointers regarding clarification about grammar and citation styles.

Acknowledgements

This thesis would not have been possible without an amazing support system that stood by me, on the days I couldn't show up for myself.

I am deeply grateful to my supervisors: Fiachra, who read through my countless senseless drafts with patience and managed to highlight ideas worth pursuing. I will always cherish our Socratic dialogues that taught me to think like a philosopher and his insistence on “crystal-clear” writing. Padraig consistently managed to uplift my spirits. I always left our meetings feeling motivated, and it's hard to emphasise how important that is for a PhD student. I will also cherish my interactions with Pat, who once told me about “running a red thread through the manuscript” in my first year. I was finally happy to find it during the final year of writing. I will be always be thankful to Ethna, my independent panel member for her kind presence.

I am extremely grateful for the support I have received from my friends and colleagues at the DCU Faculty of Humanities and Social Sciences and my community in Dublin. To Arushi, who relentlessly kept me company during Saturday writing sessions over calls—thank you for your friendship. To Aashima, who was always around and kept track of my writing schedule in the final year—I don't know what I would have done without you. Last but definitely not least, to my friends away from Dublin, thank you for the hearty conversations, motivation, and laughs over the calls.

I dedicate this work to my parents, Ashok and Manjree, who deserve every ounce of wonder and love that this world has to offer. I will be eternally grateful for the love from my sister, Ananya, and my partner, Anubhav. Their presence and quiet strength brought me back from the brink of burnout countless times. Beyond a manuscript, I owe you an enormous debt for helping me find myself during this metamorphosis. You helped me carry the weight of this journey, even when it wasn't yours to carry. And to Bholu, my golden labrador who walked beside me through earlier chapters of this life—I am certain we will meet again (probably as a cat this time).

I am grateful to my examiners, Kian and Ian, whose thorough and brilliant questions motivated me and helped me to see my work in a new light. I also extend my gratitude to Miriam for her guidance throughout the process. Lastly, I will be forever indebted to the people in academia and beyond who I have met and carried with me along the way.

Table of Contents:

List of Tables **8**

List of Figures and Graphs **9**

List of Abbreviations **10**

List of Publications **11**

Abstract **12**

Introduction **13**

Chapter 1. Climate Change and Individual Moral Obligation: A Case for Green Nudging **19**

1.1. The State of the World and Climate Change: **20**

1.2. The Ecological Systems and the Normative Turn: **24**

1.3. Features of Systemic Failure and Climate Change: **26**

1.3.1. Climate Change is in part a Behavioural Issue **26**

1.3.2. Incompatibility with Economic Growth **28**

1.3.3. Political Resistance **29**

1.4. Reasoning behind the Systemic Failure: **31**

1.5. Human Moral Obligation: **36**

1.5.1. Why do we have a moral obligation? **36**

1.5.2. The Nature of the Obligation: **39**

1.5.3. What does this obligation require us to do? **41**

1.6. Green Nudging as a Solution: **43**

1.7. Conclusion: **46**

Chapter 2. Environmental System Consequentialism **48**

2.1. Consequentialism and Climate Change: **48**

2.2. Indirect and Direct Consequentialism: **53**

2.3. Consequentialist Values: **56**

2.3.1. Hedonism **57**

2.3.2. Preference Satisfaction: **60**

2.3.3. Objective List Consequentialism **62**

2.4. Developing the theory of Environmental System Consequentialism: **64**

2.4.1. Tenets of System Consequentialism **66**

2.4.2. How System Consequentialism combines with Objective List **71**

2.4.3. How System Consequentialism combines with Ecocentrism **72**

2.4.4. A Newer Version: Environmental System Consequentialism: **75**

2.5. Conclusion: **78**

Chapter 3. Can Green Nudges help us fulfil our Moral Obligations? **80**

3.1. What is Nudging? **81**
3.2. What is Green Nudging? **85**
3.2.1. Who is a Nudgee? **88**
3.2.2. Who is a Nudger? **89**
3.3. Human Psychology, Nudging, and Moral Enhancement: **90**
3.4. How are Green Nudges Effective? **92**
3.4.1. Epistemic Myopia **93**
3.4.2. Diffused Sense of Moral Obligation **97**
3.5. Conclusion: **100**
Chapter 4. A Review of Ethical Debates on Green Nudging **101**
4.1. A Review of the Literature on Ethics of Green Nudging: **103**
4.1.1. Third-Party Harm **103**
4.1.2. Paternalism **104**
4.1.3 Green Nudges in Markets and the Role of the State: **105**
4.1.4 Sustainability goals of Green Nudges **107**
4.1.5 Autonomy **109**
4.1.6. Manipulation due to lack of Transparency **110**
4.1.7. Privacy Issues **111**
4.1.8. Fairness **112**
4.1.9. Using Nudges for long-term behaviour change **113**
4.2. How does Green Nudging raise Ethical Issues? **115**
4.3. Can we consider some ethical trade-offs if Green Nudging is promising? **120**
4.3.1. Paternalism and Green Nudging **120**
4.3.2. Fairness and Green Nudges **121**
4.3.3. Privacy and Green Nudging: **121**
4.3.4. Autonomy and Green Nudging **122**
4.4. Conclusion: **124**
Chapter 5. Green Nudging and Environmental System Consequentialism **125**
5.1. Reframing the Role of Green Nudging: **126**
5.2. Green Nudging and Systems: **130**
5.3. When are Green Nudges Ethically Justifiable?: The Rules **133**
5.3.1. Behavioural Alignment **135**
5.3.2. Awareness **137**
5.3.3. Cognitive Shifts **139**
5.4. Objective Values that Green Nudging ought to Uphold: **143**
5.4.1. Ecosystemic Objective Goods **143**
5.4.2. Objective Values for Acceptability in Human-centric Systems **145**
5.5. Conclusion: **150**
Chapter 6. An Account of Trust in Green Nudges **151**
6.1. The Basis for Trust in Green Nudges: **153**
6.1.1 The Need for Trust: **153**

6.1.2. Trust and principled autonomy: 155
6.1.3. Practical Advantages of Trust for Effectiveness, Prudence, and Fairness:. 157
6.2. Trust and Trustworthiness: A Theoretical Framework: 159
6.2.1. Trustworthiness: 161
6.2.2. Normative basis for trust: 165
6.3. Trust in Green Nudges: 168
6.3.1. Nudger demonstrates trustworthiness: 168
6.3.2. Nudgee assesses trustworthiness: 182
6.4. Conclusion: 185
Chapter 7. Behavioural Self-Governance and Other Recommendations for Green Nudging 187
7.1. The Need for Procedural Recommendations: 187
7.2. Honesty and Accessibility of Green Nudges: 190
7.3. Minimising the scope of Consumptive Nudging: 192
7.4. Behavioural Self-Governance and Green Nudges: 195
7.4.1. Self-Governance: 198
7.4.2. Stakeholders in Behavioural Self-Governance: 203
7.4.3. Design Principles of Behavioural Self-Governance: 207
7.5. Closing Remarks: 213
Chapter 8. Conclusion 215
8.1. Overview of the work: 215
8.2. Broad Reflections: 218
8.3. Contribution and Future Directions: 219
8.4. Limitations: 221
8.5. Final Thoughts: 223
References: 225
Appendix A: Systematic Methodology: 244
Appendix B: Design Principles by Ostrom (1999): 250
Appendix C: Design Principles in Co-Cities Project (2019): 251

List of Tables

Table 1: Stakeholders in Behavioural Self-Governance……..…..…206

Table 2: Details of Search Strings and Results……...……...245

Table 3: Inclusion and Exclusion Criteria……..…..……..…..…247

List of Figures and Graphs

Figure 1: The Pro tanto Model by DesRoches et al. (2023)…….…...…...116

Figure 2: A four-step PRISMA-Flow diagram of the literature search……..…249

Graph 1: Distribution of results into different categories during the screening process…248

List of Abbreviations

GN…..Green Nudging

ESL….Environmental System Consequentialism

BSG….Behavioural Self-Governance

List of Publications

Kuyer, P., Shukla, S., O'Brolcháin, F., Rodríguez Doncel, V., & Gordijn, B. (2025). The ethics of digital nudging for sustainable energy consumption – A closer look at the EU Green Deal. *Technology in Society, 83,* 102990. https://doi.org/10.1016/j.techsoc.2025.102990

Shukla, S., O'Brolcháin, F., Murphy, P. (manuscript in revision). Ethics of Green Nudges–A Systematic Literature Review. *Humanities and Social Sciences Communications*.

Abstract

Ethics of Green Nudging

-Sukanya Shukla

This research positions green nudging (GN) as a case study to conduct an ethical analysis through systems: locating the ethics of using GN at the intersection of environmental systems, cognitive systems, and human-centric systems, such as socio-political systems and technological systems. GN is a strategy that leverages psychological features, such as self-image, inertia, and social image, to promote pro-environmental behaviour.

Despite the widespread application of nudging, the ethics of GN remain underdeveloped. This study explores literature on human moral psychology to illustrate the effectiveness of GN. It also identifies ethical issues associated with its use. To evaluate the ethical permissibility of GN, it adopts a systems-based lens to examine the role of green nudges in light of the climate crisis. To this end, this dissertation develops a novel philosophical approach, *Environmental System Consequentialism*. To further the ethical examination of GN, this work takes a pragmatic turn to propose a morally justifiable account of trust. Trust is illustrated as a functional condition for green nudges to be ethical and effective.

To begin, it is shown how GN can address the effects of psychological traits that prevent one from fulfilling their environmental obligations. Next, ethical issues and themes emerging from the use of GN is examined and identified through a literature review. Then, the theoretical apparatus developed in the dissertation provides a coherent consequentialist position to examine the ethical permissibility of green nudges. Next, the role of institutional trust between the individuals being nudged (nudgee) and the institutions implementing GN (nudger) is explicated for green nudges to be effective, as well as ethical. The signifiers of trustworthiness and the conditions for trust are sketched. Building on this analysis, the dissertation concludes with procedural recommendations for ethical implementation and proposes Behavioural Self-Governance (BSG).

Introduction

Climate change is a complex problem that, through the deep embeddedness of human-made systems within ecological systems, has revealed humanity's psychological and moral shortcomings (Kaaronen, 2018). Our behavioural responses, which shape these human-made systems, are driven by insatiable hunger for resources, power and domination (Albert, 2022). These inadequacies scaffold the larger socio-economic and political systems that continue to fail in responding to this crisis (Gupta et al., 2024; Albert, 2022; Aligica & Boetke, 2009). Some insist that incremental reforms or rational appeals may help us navigate this collective challenge; nonetheless, history shows that such hopes remain unrealised (Gardiner, 2011; Nielsen et al., 2021). Humans, acting and contributing across numerous systems, continue to fuel this crisis even though we stand at the precipice of a catastrophe, with evidence about future perils (IPCC, 2023; Savulescu & Persson, 2017a). Ultimately, even though we have become more technologically advanced, our collective response remains unsatisfactory and short-sighted (Savulescu & Persson, 2017a; Singer, 1972). This raises the question about what potential changes in human psychology and morality, perhaps driven by technological advancements, can respond to this crisis in a timely and effective manner.

Humanity has seen immense progress in technology and science over the past few centuries. We invented steam engines and aeroplanes, and now stand closer to developing Artificial General Intelligence (AGI), a possibility only dreamt of a few decades ago (Bubeck et al., 2023). Yet despite this technological progress, some argue that we remain unprepared to confront the collective and existential threats of our age (Savulescu & Persson, 2017a; Savulescu & Persson, 2012). This is surprising considering how we have become more informed about the world through scientific discoveries and emerging technologies. However, this progress has not translated into sufficient action in light of climate change

(IPCC, 2023). This dilemma reflects an important insight about human nature: we possess certain epistemic and moral limits (Persson & Savulescu, 2015; Savulescu & Persson, 2012). Thus, the challenge to prepare human nature to deal with climate change through technology becomes a pragmatic as well as an ethical challenge.

Philosophers have, for centuries, argued for reflection and critical engagement in the face of problems– from Plato's allegory of the cave to Phillipa Foot's trolley problem. Environmental thinkers have considered climate change as the central problem of our age and have demanded a reformation of human morality–more specifically, they have appealed for a transformation in how humans understand their relationship with nature (Gardiner, 2011; Naess, 2005; Leopold 1949). Yet despite law and politics often looking up to ethics for guidance, the ideas of environmental philosophers have struggled to find mainstream applications in environmental policy and governance (Gardiner, 2011).

On the one hand, accelerating technological progress to address climate change is evident through autonomous devices, Internet of Things (IoT), and energy-related advancements like carbon-capturing, smart grids, and solar-powered vehicles. On the other hand, we tend to overestimate that these technologies can rescue us without a fundamental change in human psychology and moral disposition (Persson & Savulescu, 2015). The difficulty lies in the unpalatability of human-centric systems, values, and norms to integrate with the insights from environmental ethics (Kaaronen, 2018). This invites us to reflect on whether changing/improving human psychology, perhaps with the help of technology, can help us become better moral agents in light of climate change. Moreover, there also remains the question whether human-centric systems are willing to accept this transformative proposal about human psychology.

This thesis attempts to explore these tensions between climate change as a systemic issue, human-centric systems, environmental systems, technological systems, and human

cognitive systems. In doing so, it takes a different approach–it engages in an ethical inquiry by exploring the role of these systems and their interactions with one another in light of climate change. The case of Green Nudging (GN) provides an opportunity to explore these tensions. It also provides an opportunity to integrate human moral psychology with technological tools to help us mitigate the daunting challenge of climate change. Human moral psychology refers to the view that ethical decision-making originates in psychological systems (Tjeltveit, 2021).

GN stands at the intersection of different systems. In looking at the ethics of GN, this thesis develops the theoretical apparatus of *Environmental System Consequentialism* (ESC) to help us think through systems. This ethical theory focuses on accommodating the complex nature of climate change as a systemic issue with interconnected systems–the ecological systems at stake, but also the social and political systems that affect it and depend on it. This pluralistic consequentialist theory argues that an action is justified if it accords with the rules that generally maximise the well-being of ecological systems, but also upholds certain objective systemic values essential to other systems (as well as ecological systems).

GN can change human behaviour, perhaps in an advanced form; it can lead to long-term changes in human psychology to deal with the climate crisis (Kaaronen, 2018). GN emerges from advancements in psychology and neurosciences; we now know more about thumb-rules (or heuristics) and traits inherent to human psychology (Schubert, 2017b). We are finding ways to integrate with technological advances to learn more about human behaviour: we have smart devices attached to our bodies that can collect, organise, and process data about our decision-making patterns (Ranchordás, 2020).

While it is true that technological progress will improve the performance of GN and similar strategies over time, technological systems in the current age have their own perils (Ranchordás, 2020). The questions about human autonomy and paternalism remain, as there

remains a risk of informational privacy, fragmentation of preferences, and manipulation at the hands of those who can misuse or misapprehend the technologies that were designed to guide us out of the crisis of our times (Hilton et al., 2018; Schubert, 2017b; Bovens, 2009). The difficulty lies in finding ways to integrate what philosophers, especially environmental philosophers, have long argued for: a transformation of human relationships with nature (Kaaronen, 2018). We are then compelled to reflect upon ways in which we can leverage technological progress to manifest a better human-nature system, while avoiding the pitfalls of modern scientific/technological progress. This leads us to the main predicament that this dissertation seeks to address: *The Ethics of Green Nudging.*

The ability to nudge may be effective for changing human psychology, but it has its own perils in the wrong or incompetent hands (Lenzi, 2019). The proposal of integrating technological prowess to tackle a climate crisis that is so impervious to human psychology and morality is not possible without exploring pragmatic pathways (Persson & Savulescu, 2015). This insight is explored in light of these tensions between human-made social, political, technological systems, environmental systems, and the human cognitive systems. In doing so, this work also attempts to analyse factors that can make the systems of institutional relationships better. This is where trust between institutions and individuals (people being nudged) becomes central. While trust is ubiquitous, it is also fragile (Baier, 1986). Moreover, given that the technological levers may be in the hands of powers that can misapprehend or misuse trust, it is vital that we trust only those who are trustworthy (O'Neill, 2002a). Thus, this pragmatic turn in this work, along with procedural recommendations to legitimise GN, provides pathways for exploring trust in other technological solutions, which can guide us through the modern issues of our time.

The reader is encouraged to keep a braided approach as they reflect on these tensions, objections, and solutions. To keep a braided approach is to see this ethical analysis of GN as

an exemplification of ESC, while also considering the climate crisis and the potential technofixes that ESC may inform. Some may be tempted to see GN as a straightforward technological solution, yet it remains tethered to the perils of modern technology and the evolutionary cognitive systems. This is to say that we ought to see GN in light of the complex systems within which it operates. At first glance, this encourages the reader to consider the potential of GN and ethical pathways to implement it. But broadly, the reader is invited to keep applying systems-based thinking to tackle the problem of climate change. Taking a systemic turn in ethical examination is to think through systems on every level: to see climate change as a failure of political, moral and social systems, to see GN as a system between the institution implementing nudging (nudger) and the individual being nudged (nudgee), and to understand that whatever we do is not independent of a system.

To this end, this dissertation unfolds in three arcs. The first arc examines climate change as a systemic moral challenge and develops *Environmental System Consequentialism* (ESC) as a framework for ethical reasoning in light of climate change (Chapters 1-2). The second arc explores green nudging, its effectiveness in shaping behaviour, and the ethical issues it raises; and finally defends green nudges from an ESC approach by showing how they are in accord with the rules that maximise good ecosystemic outcomes and illustrating the objective value list that green nudges ought to uphold/preserve to be ethically justified (Chapters 3-5). The third arc explores trust as a functional value for an ethical system between the nudgee and the nudger and offers recommendations for ethical integration of GN into political and economic systems, before concluding with the broader implications of this work (Chapters 6-8).

This research invites readers to see the complexity of moral decisions through systems and make choices while acknowledging it. Some may be tempted to treat ethics as a matter of isolated principled judgement, yet our actions pull a number of strings across different

systems, so the consequences of those actions cannot be judged in isolation. Ethical judgements, therefore, ought to be made while considering the complexity of the systems rather than in isolation. Looking at applied ethics in this fashion has interesting implications. The emergence of systems-based thinking in modern sciences, from ecological sciences to urban geography, has provided us with the opportunity to understand the world and the consequences of our actions in light of systems (Hiller, 2016). It captures the complexity of judgements and also the dilemma of choosing between different values and stakeholders. Bringing this approach to ethics is relevant to the problems of modern times: an appeal that echoes the calls of environmental philosophers for moral transformation, behavioural experts who explore the limits of our cognitive systems, and political theorists who remain perplexed by the complexity of social, economic, and political systems (Kaaronen, 2018; Gardiner, 2011).

Chapter 1. Climate Change and Individual Moral Obligation: A Case for Green Nudging

The human impact on Earth systems is immensely serious and requires an urgent response (IPCC, 2023; Garvey, 2011). Indeed, the climate crisis poses a catastrophic threat which may lead to a systemic collapse of planetary boundaries and ecosystems[1] (IPCC, 2023; IPCC, 2022). The scientific consensus is that significant behavioural shifts on individual and collective levels are needed (amongst other systemic changes) (Nielsen et al., 2024; IPCC, 2023; Nielsen et al., 2021b; IPBES, 2019). However, as we will see as a recurring theme in this chapter and Chapter 3, human behaviour in response to climate change continues to fall short in mitigating this problem. Such a failure is rooted in our evolutionary inertia[2] that underlies climate-harming actions.

Given the vast potential for harm arising from climate change and other environmental catastrophes, we, as humans, have at least a collective practical interest in minimising the impact of the damage. Indeed, this thesis shows that we have a moral obligation to mitigate the most deleterious consequences of climate change. Nonetheless, we are not taking the necessary measures to meet these practical interests or moral obligations.

This lack of action to address the climate crisis has led many to call for radical changes, ranging from technological revolutions (such as geoengineering and human enhancement) to moral or political transformations (including an end to economic growth or the expansion of

[1] The implications of the extreme events (extreme weather, droughts, cyclones, forest fires) on Earth's systems clearly signal a catastrophic change and an existential threat to the human population, especially for those residing in climate-vulnerable areas.
[2] Chapter 3 illustrates this point by showing how green nudges help individuals act despite the evolutionary psychological traits that hinder pro-environmental actions.

the moral circle to include non-human entities) (Gupta et al., 2024; Persson & Savulescu, 2015; Singer, 1981).

This thesis explores the potential of "Green Nudging" to help us improve our response to the global environmental crisis. Nudging, as we will see, is a tool that can subtly alter behaviours and influence the patterns, but it also brings with it moral dilemmas. This thesis will explore whether nudging people for the sake of pro-environmental outcomes–green nudging–can be ethically justified.

1.1. The State of the World and Climate Change:

The latest report from the Intergovernmental Panel on Climate Change (IPCC), the leading source of scientific information on the status of climate change, defines climate change as a rise in global temperatures unequivocally driven by human activities (IPCC, 2023, p. 42). An increase in global temperatures destabilises various ecosystems, including the oceans, freshwater bodies, forests, and cryospheric systems, disrupting their ecological functions and causing substantial damage that is often irreversible (IPCC, 2023).

As ecosystemic balance[3] and human vulnerability are interdependent, climate change has indisputable implications for the present and the future well-being of human life on the planet (IPCC, 2022; Albert, 2022; IPCC, 2023).

The mean increase in global temperatures of Earth's atmospheric systems (in 2020) has been observed to be 1.1°C (or degrees Celsius) above pre-industrial levels, referred to as the period between 1850 and 1900. The report states with high confidence that surface

[3] Ecosystemic balance essentially refers to the well-being of the Earth's ecosystems broadly, as well as the interconnectedness and interdependencies existing between all other smaller ecosystems, which are subcomponents of the planetary biosphere. To be healthy, these smaller ecosystems, ranging from forests to deserts, should be in a state of dynamic equilibrium. In other words, the well-being of one system has an impact on others. Disruptions in the ecosystemic balance caused by human-induced activities can trigger effects that compromise the well-being of the components of the ecosystems, such as humans and non-humans, as well as their ability to self-heal after disruptions (IPBES, 2019).

temperatures have been increasing at a faster rate since 1970, driven by human activities, compared to any fifty-year time frame in the past 2,000 years (IPCC, 2023). This trend emphasises the role of human activities as the main drivers of climate change, with their contribution to the issue becoming increasingly pervasive in recent decades.

More than 99 per cent of peer-reviewed published scientific literature suggests that climate change is an anthropogenic (human-induced) issue (Lynas et al., 2021). This supports the evidence that human activities, such as the use of fossil fuels, industrial emissions, land and forest use, contribute to climate change. Over the centuries, humans have prioritised rampant economic development by using up natural resources such as oil, coal, natural gas, and timber (Singer, 2011). The extraction, processing, and use of these resources emit greenhouse gases (GHGs), such as carbon dioxide, hydrofluorocarbons, and methane, into the Earth's atmosphere. The atmosphere, as a blanket that maintains the Earth's temperature, traps GHGs, which increases surface temperatures, causing climate change (IPCC, 2023; IPCC, 2022).

The current warming levels have already exceeded 1.5 °C, as reported by the World Meteorological Organisation[4] (WMO) in 2024. This suggests that we may have already missed significant opportunities to address climate change[5], while persistent threats to human well-being, biodiversity, and agroecosystems continue to escalate. Surpassing the 2°C threshold could trigger severe and irreversible consequences for both human societies and

[4] World Meteorological Organisation (WMO) is a specialised agency of the United Nations for providing scientifically backed information about climate, weather, and water resources related support to its members.
[5] The Paris Agreement aims to limit long-term global warming to well below 2 °C, with efforts to limit the rise at 1.5 °C. The Agreement to prevent warming over the long term between 1.5 °C and 2 °C has not been violated yet. The goal is to limit the increase "well-below 2 °C". Climate models suggest that we have likely entered a 20-year period in which the 1.5 °C average threshold will be crossed. Considering the current trends, the long-term 1.5 °C limit is most likely to be crossed in the late 2020s or early 2030s.

ecosystems, including lethal heatwaves, Arctic ice melt, rising sea levels, and the extinction of many plant and animal species (United Nations Climate Summit, 2023).

The consequences of rising surface temperatures are alarming. The IPCC report (2023) suggests, with high confidence, that there is evidence that human influence on environmental systems has increased the likelihood of compound events, such as droughts, heatwaves, precipitation, flooding, and tropical cyclones. These localised catastrophic events cause unprecedented harm to both human and non-human beings. Moreover, since the local ecosystems across the planet are interconnected and interdependent, such impacts can cascade into multiple stressors. For instance, a shift in rainfall patterns can affect other ecosystem services, such as land management or fisheries, subsequently leading to widespread changes in socio-economic systems and human livelihoods (Lawrence et al., 2020). There also remains a possibility of major existential threats resulting from cascading effects due to a rise in global surface temperatures, a scenario described to be grim by Integrated Assessment Models (IAMs) (Appleton, 2008).

This has implications for human lives and well-being. For instance, rapid flooding in the low-lying plains of Pakistan threatens millions of people with displacement and water-borne diseases, and prolonged droughts result in starvation, conflict, and insecurity for people in the African sub-continent (Sajjad & Khan, 2025; Mishra, 2025). Moreover, vulnerable communities, which have historically contributed less to climate change, are more at risk (IPCC, 2023; IPCC, 2022; Tol, 2009).

At least 3.3 to 3.6 billion people worldwide live in areas highly vulnerable to climate change (IPCC, 2023, p. 51). Tropical countries like India are already dealing with severe climate impacts—over 64 per cent of its districts have experienced increasingly intense monsoon seasons, resulting in widespread flooding, destruction of homes, and agricultural losses compared to the 1982-2011 baselines (Council on Energy, Environment and Water,

2022). Moreover, heat-related productivity losses are projected to result in the equivalent of 80 million full-time job losses globally by 2030, a trend which is already visible in India, where rising temperatures are leading to migration of rural agricultural workers toward urban areas in search of more stable livelihoods (Kuhn, 2025). Such disruptions in the Earth's ecosystems are already causing, and will continue to cause, physical and mental suffering to the global population, unless radical steps to tackle climate change are taken (IPCC, 2023; Padhy et al., 2015; Tol, 2009; Haines & Patz, 2004).

The impact of climate change is experienced not just by the human population but also causes the loss of biodiversity across the planet. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) provides information and policy recommendations on climate-related biodiversity loss and the impact of climate change on human-nature interactions. By 2016, the report notes that more than 6,000 domesticated animal breeds used for food production had become extinct, and approximately 1,000 more species were threatened (IPBES, 2019). This is because nature[6] has been significantly altered due to the effects of climate change. A warming of 1.5 °C to 2.0 °C would dramatically reduce the possibility of survival for countless species, especially those that are already vulnerable due to habitat loss, pollution, and the overexploitation of agroecosystems (IPBPES, 2019).

As previously mentioned, the ecosystemic balance and vulnerability of living beings on the planet are interdependent (IPCC, 2022; Albert, 2022; IPCC, 2023). To put it another way, climate change does not occur in isolation. The effects of changing global surface temperatures, or the climate crisis, are felt and experienced across decades and are difficult to predict (IPCC, 2023; Nefsky, 2021; Garvey, 2011).

[6] The report defines the term 'nature' as embodying 'different concepts for different people, including biodiversity, ecosystems, Mother Earth, systems of life and other analogous concepts'. (IBPES, 2019, p.10).

The effects are interlinked, cascading, and slow-moving[7] (Gardiner, 2011). Not only are local ecosystems, such as forests, rivers, and plains, affected by the impact of climate change, but also the larger ecosystem that comprises them all, which is the Earth system. Therefore, understanding the nature of the climate crisis requires considering how it affects the beings on the planet, as well as their causal relationships and the correlations between them, together comprising the Earth System.

Recognising this complex characteristic of the climate change problem, this thesis will argue for an ethical theory that values systems and their relationships, in Chapter 2. The following section outlines the nature of the issue, which is systemic and also complex. This complexity is best understood by examining how different systems created by humans and centred on humans, such as social and political systems, play a pivotal role in undermining efforts to address this crisis.

1.2. The Ecological Systems and the Normative Turn:

The Earth can be characterised as a system—an intertwined and interconnected set of biophysical subsystems (such as biosphere, hydrosphere, cryosphere, lithosphere). Within this broader system, the ecological systems, the biosphere and its interactions, incorporate the system that sustains life and enables flourishing of both humans and non-humans. This is the system on which all known life depends.

The Earth system has been co-evolving with human social systems (economies, technologies, cultures), which are dependent on the ecological system but also have an impact on its well-being. This is empirically the case and is also morally important, as all

[7] To say they are slow-moving means that the visible consequences of climate change, for humans, can take months and often decades to be visible. For instance, the melting of icebergs causes events such as changes in rainfall patterns and flooding. This process is slow and not easily comprehended by humans, who are evolutionarily more equipped to deal with immediate dangers. Chapter 3 expands on these limitations about human nature.

systems of value are dependent on the functioning of the ecological systems. This includes human moral, political, economic, and social systems that are embedded within the Earth system or the global environment.

Human flourishing, to a large degree, depends on maintaining the health of the ecological system, which is only possible if the Earth remains hospitable to humans. Climate change threatens this possibility by rendering large parts of the planet uninhabitable, or at the very least by creating enormous hardships, making life for future generations much more difficult. As such, maintaining ecological systems is not only a practical concern but also a moral concern.

Anything else that humans value–freedom, wealth, enlightenment, happiness–is at least somewhat dependent on the functioning of the ecological systems. Thus, as far as this thesis is concerned, ecological systems that are pivotal for maintaining the Earth's systems are the systems of moral value. Nonetheless, action to keep ecological systems has been unsatisfactory.

The fact that any ethical work focused on environmental well-being must take systems into account suggests that a system consequentialist position[8] provides a helpful position. In short, in examining the climate crisis, we should focus our ethical lens on the consequences for the systems that support life.

Furthermore, the system consequentialism that this thesis is concerned with also considers systems to have some objective goods, such as flourishing ecosystems, a stable climate, and healthy humans and biodiversity. These objective goods are essential for the pursuit of any other (objective) human goods, e.g. happiness, liberty, justice, etc. This

[8] A consequentialist theory of systems is formulated by Avram Hiller (2011, 2013). He focuses on developing system consequentialism as a fully consequentialist theory in environmental ethics. The tenets of this theory are discussed in detail in Chapter 2.

approach is *system consequentialism*, which holds that actions are right to the degree that they lead to good outcomes for the ecological systems[9] (the systems that make possible objective human and non-human flourishing).

1.3. Features of Systemic Failure and Climate Change:

Whilst the scientific evidence that climate change causes unprecedented and morally concerning consequences is clear (as discussed), humanity's collective response has been insufficient (Albert, 2022; Garvey, 2011). It is reasonable to wonder why so little has been done.[10] As the problem of climate change is systemic in nature, various human-created systems, including social and political systems, play a role. Moreover, behavioural systems on both individual and collective scales, also influenced by existing political and economic mechanisms[11], play a role in our inadequacy to address this issue.

1.3.1. Climate Change is in part a Behavioural Issue

While the need for drastic political and economic measures to prevent such a global catastrophe cannot be overlooked, it is also important to recognise that the root of this failure lies in human behaviour, maladapted to address climate change. In other words, the factors that drive human behaviour are not well-adapted to prevent rapid global surface warming.

Savulescu and Persson (2015) also note, "One central, neglected problem is that a significant cause of these problems [climate change] is human behaviour,… adapted to life in small, close-knit societies…" (p.338). This is to say that climate change is largely a behavioural issue as well. Thus, to understand the lack of collective response to prevent

---

[9] Chapter 2 formulates this philosophical position, which then informs an ethical defence of Green Nudging (GN), the focus of this study in Chapter 5.

[10] It is also possible that there is a deeper, conceptual problem that distorts how we perceive the world, e.g. an anthropocentric perspective that only places moral value on human benefits underpins our global institutions and collective and individual behaviours. To address climate change effectively, we must understand the psychological barriers preventing us from taking action, a topic explored in Chapter 3.

[11] Some views question the validity of the scientific evidence behind climate change (or climate change denialists), which also hinder effective action in this context.

climate change, we must focus on the underlying behavioural factors that continue to hold us back from mitigating climate change.

Evolution has shaped our behaviour. Certain behaviours, such as caring for groups, ensuring all group members are treated fairly, and providing food and shelter to sick members whenever needed, have helped humans survive over time. While humans adapted these behaviours through evolution, they continue to shape our responses to the climate crisis today[12] (Merz et al., 2023; Singer, 1981).

Due to centuries of GHG emissions, we are now proceeding towards a full-blown crisis, while our impulses continue to be shaped by maladaptive behaviour. Our current psychological dispositions hinder our ability to address climate change, which requires us to prioritise long-term goals. These psychological traits are a proof that, despite the advancement of technology and science, our minds have not evolved sufficiently to comprehend the complex failures caused by climate change (Kaaronen, 2018). Such an evolutionary inertia is characterised by human tendencies adapted during evolutionary phases such as watching out for immediate attacks, short-term gratification, and self-interest that were adapted during evolutionary phases (Persson & Savulescu, 2017b, 2015). For instance, nations continue to invest in nuclear-weapon programs, arms, military, and missiles because they prepare them for immediate threats, but fail to show vigilance about the danger of climate change. Similarly, individuals from higher socio-economic groups continue to take private flights every other day without recognising the environmental impact resulting solely from their transportation choices.

[12] For instance, evolution shaped us to stay in groups to survive immediate dangers, accumulate resources to ensure survival, and prioritise procreation to continue the genetic lineage. These tendencies continue in today's age, where we fail to psychologically grasp the severity of climate change and continue to amass wealth while thinking of the future benefits.

Apart from the complex social and political systems that are not well-adapted to address climate change, on an individual level, we continue to function as 'self-interested utility maximisers' (Merz, 2023, p. 8). Currently, an average American citizen uses 175 times more fossil fuel than their counterpart in the 1800s, and the statistics are likely to worsen with the growing global population (Merz, 2023; Nolt, 2011).

Apart from psychological factors driving behaviour, one of the reasons why human behaviour fails to adapt is the excessive focus on economic growth. For instance, India's population continues to grow, with near-term ambitions to achieve rapid economic growth. This suggests that an increase in population will lead to a significant rise in resource consumption and waste production. Moreover, such a growing economy would also be characterised by an average citizen striving to climb the socioeconomic ladder, leading to increased conspicuous consumption. The next subsection discusses this issue.

1.3.2. Incompatibility with Economic Growth:

Another reason why we have failed to address climate change so far is the persistent prioritisation of economic systems by humans, often at the expense of environmental harm. This prioritisation partly results from what Michael Albert (2022) calls a "continuationist" bias, an assumption that the existing model of economic growth can continue in the future without inhibiting the efforts to address climate change on a global scale.

Such a bias undermines climate action. For instance, over the past five decades, the global population has doubled, while the world economy has grown four times faster, resulting in increased demand for energy and materials to meet the growing population's needs (IPBPES, 2019, pp. 13-14). Most nations, including both the least developed countries and developing countries with growing populations, have ambitions to expand their economic activities, often at the expense of environmental harm (Gupta et al., 2024; IPBES, 2019).

Given this context, it is not possible to achieve the climate targets while continuing the current trajectory of economic growth (Albert, 2022). Economic growth is driven by and in turn increases production and consumption, contributing to continuous environmental degradation (Kirby, 2024; Albert, 2022). Such a continuationist assumption of economic growth focuses on resource-intensive production to meet demands and favours fossil fuel consumption, and thus, is incompatible with addressing climate change. Therefore, while the urgency of addressing climate change demands systemic transformations, the current economic systems resist such change[13].

One of the possibilities is to rapidly replace fossil fuels that cause GHG emissions with renewable or alternative energy sources, a hope found in almost all IPCC reports so far. Another possibility is to focus on the 'Negative Emissions Technologies' (NETs), such as Direct Air Capture, afforestation and reforestation, and regenerative agriculture (IPCC, 2023; Albert, 2022). These technologies and alternative energy sources, however, can achieve the climate targets only if the use of fossil fuels is phased out substantially (Gupta et al., 2024). An underlying difficulty in adopting such pathways is that there remains political resistance to implementing the necessary economic changes to prevent global warming from exceeding 2 °C.

1.3.3. Political Resistance:

The existing political systems also hinder the pursuit of climate targets. There are many vested interests that will be disadvantaged, such as oil-rich states, intensive agriculture, the fossil fuel industry, fishing, mining, and logging. For instance, public choice theorists,

[13] While strategies like carbon pricing address misaligned incentives by internalising environmental costs, they function through economic mechanisms and do not directly engage with the psychological and behavioural dispositions that underlie climate-harming behaviour. GN complements such strategies by targeting the cognitive and motivational dimensions of behaviour that economic incentives cannot reach alone (Gravert & Shreedhar, 2022).

such as James Buchanan, contend that political institutions are driven by self-interested actors whose decisions often diverge from the ideals of public good (Aligica & Boetke, 2009).

When small, organised groups benefit from specific policies, they have strong incentives to engage in rent-seeking, that is, using their power to extract benefits while imposing costs on others (Aligica & Boetke, 2009). In doing so, these groups exert influence over both local and global politics (Gupta et al., 2024; Carter, 2025). While there is lobbying on regional levels to influence policies, there is also a lack of political action on an international scale. Fossil-fuel giants also continue to enjoy an 'outsized influence' in international negotiations on the limits on GHG emissions and climate treaties, namely, UNFCCC talks on Corporate Accountability and COP talks, including the upcoming one in Brazil (Harvey, 2025; Gupta et al., 2024).

In summary, the status quo human behaviour and human-constructed economic and political systems reveal the complexity of climate change as a systemic issue. Their relationship with the environment and influence on one another continue to harm the ecological systems, systems that are critical for the well-being of life on this planet.

Moreover, this systemic failure is on a collective level. Climate change can be viewed as a "collective action problem" because there are billions of people and systems driven by humans that need to cooperate and work in conjunction with one another to tackle this issue substantially. This means that society can only benefit if individuals and groups cooperate, both now and across generations.

This fact is particularly difficult for humans to confront. We can cause catastrophic harm unintentionally, through selfishness, or through a lack of constraint regarding spending, consumerism, political power, and industrial production. There are also evolutionary factors underlying this lack of comprehension (discussed in Chapter 3). As a result of these factors, the economic and technological systems we have built, which have global and

intergenerational impacts, are anthropocentric and blind to how they stretch the planetary limits and degrade the ecological systems. The following subsection critiques this anthropocentric line of thinking and urges us to consider a non-anthropocentric stance.

1.4. Reasoning behind the Systemic Failure:

An anthropocentric perspective on the world has informed human response to climate change. A commitment to anthropocentrism holds that our moral obligations exist for the sake of human well-being (Des Jardins, 2001). Environmental philosophers argue that a human-centric ethics is an inconsistent moral position. They find no justifiable reasoning for holding humans as more valuable than other parts of nature. For instance, environmental utilitarian argue that many animals can experience pleasure and pain just like humans. Thus, if pain and pleasure are morally significant experiences, there is no reason why moral value cannot be extended to non-human beings subjected to pain because of environmental degradation.

It has been argued that because humanity places no moral value on animals, the biosphere, or ecosystems, we do not care when we harm them. This lack of moral value means we don't care for them; we are blind to the damage we do to the environment or its constituent members. Such a view is anthropocentric, which underlies the systemic issue of climate change.

As anthropocentric approaches have failed to help us act decisively on this collective crisis, some environmental philosophers argue for a non-anthropocentric worldview. They call for approaches that value other living beings and flourishing systems. Non-anthropocentric philosophers appeal to us to recognise the 'weakness of current orthodoxy,' which views conventional human good as valuable, thereby ignoring the interests and preferences of beings other than humans (Gardiner, 2011, p. 220). A non-anthropocentric

position extends our moral obligation beyond human beings. It urges us to consider other components of the natural world as morally valuable.

However, even within non-anthropocentric views, there are different schools of thought. Non-anthropocentric views, such as those held by Peter Singer, Robin Attfield, and Paul Taylor, provide different reasons for holding the view that non-human beings are worthy of moral consideration.

As a preference utilitarian, Peter Singer argues that, given that non-human animals are capable of feeling pleasure and pain, they have preferences for pleasure and avoidance of pain. Singer's view is *sentientism*, developed in *Animal Liberation* (1975) and several other key texts. From this view, it is ethically impermissible to subject sentient beings to suffering by consuming their meat, keeping them in small, cramped cages, giving them electric shocks, and administering them with needles to test medicines and cosmetics for human benefits (Singer, 1975). Moreover, as emissions rise, global surface temperatures damage the ecosystems, such as forests, lakes, and rivers, on which these sentient beings rely for survival; our actions also impose harm on them. Singer's (1975) view holds that obligations to address climate change exist because subjecting sentient beings to pain and suffering is morally wrong.

Robin Attfield's (2003, 2012) non-anthropocentric view is *inegalitarian biocentrism.*[14] His view differs from Singer's (1975) position on the grounds that all living beings, including non-sentient and sentient beings, experience health, disease, and injury[15].

[14] Attfield's (2003, 2012) inegalitarian biocentrism posits that all living beings possess intrinsic value, but as an inegalitarian, he argues that different living things have varying moral worth.

[15] Sentientism considers the ability to feel pleasure and pain as the basis for moral worth. This account values non-sentient beings purely on instrumental grounds (Singer, 1975). This means that Singer's position holds that we must value other non-sentient beings only to the extent their well-being is valuable for the sentient beings. While Singer (1975) does not hold his view to be inegalitarian, differentiating between moral consideration based on sentientism brings his position closer to Attfield's (2003). This poses a problem for these two non-anthropocentric views. For instance, in theory, both of these views can allow large-scale destruction of corals because they are seen as non-sentient to save a human who is sentient (according to Singer (1975)) or has a greater moral value according to biocentric inegalitarianism.

His account considers it morally valuable to prevent harm that compromises the well-being of other living beings, including humans.

He writes, "...since beneficence is central to ethics, we should avoid inflicting harm and injury where there is no good reason to do so; in other words, living creatures matter even if they are not sentient" (Attfield, 2012, p.85). The core tenet of Attfield's biocentrism inegalitarianism suggests that all living creatures have an intrinsic value; their life, conditions for flourishing, and attaining their good are intrinsically valuable (Attfield, 2012). That is, all living beings warrant moral consideration for their own sake and not on purely instrumental grounds.

In contrast to Attfield, who focuses on the conditions for flourishing, Paul Taylor (1981) adopts an egalitarian position, arguing that a specific moral attitude, 'respect for nature,' serves as the central foundation of environmental ethics. He argues that we must recognise that each living thing, whether sentient or non-sentient, has a good of its own; recognising this and pursuing this intrinsic good is to revere 'respect for nature' (Taylor, 1981). He writes, "To put it in a Kantian way, to adopt the attitude of respect for nature is to take a stance that one wills it to be a universal law for all rational beings." (Taylor, 1981, p.05).

In being an egalitarian, as opposed to Attfield (2003), Taylor[16] (1981) holds that this tenet (respect for nature) requires us to hold every living being as equally worthy of respect, and one must act to uphold their good without exception. Such an obligation requires us to

[16] Taylor's (1981) biocentric egalitarianism reveres 'respect for nature'. Such an environmental ethic recognises that all living beings are "teleological centers of life" (Taylor, 1981, p.09). This is an Aristotelian influence on Taylor's work, where the Greek word *telos* translates to purpose, goal, or end (Des Jardins, 2001). The core idea of Taylor's ethic is that every living being has an inherent worth because it has an objective good of its own, or a *telos*, regardless of whether it is consciously aware of it. Moreover, humans are not inherently superior to other beings (Taylor, 1981). Based on these two premises —the inherent worth of every being and the rejection of human beings as a superior species —Taylor (1981) argues for an environmental ethic that grants equal moral standing to all living beings. Taylor (1981) focuses on the moral status of each living organism rather than ecosystems. This means that his account excludes non-living entities like rocks, soil, and sand, unless their well-being contributes to the "ecological niche or function" of living beings (Des Jardins, 2001, p. 141).

consider the interests of all affected beings, resulting from a universal duty to act with reverence for nature.

All these non-anthropocentric positions, sentientism, inegalitarian biocentrism, and egalitarian biocentrism, have different reasons for arguing that a moral obligation to preserve the environment exists. However, all of these positions are individualistic in nature. Climate change is a systemic issue, as shown. To capture the systemic nature of the crisis, we require an ethical theory that considers systems as morally valuable.

These positions discussed above ignore an essential aspect: the relationship between different components of the system. This is specifically important concerning climate change, where human influence trickles down the various components of the ecosystems to result in the catastrophic effects of climate change. This influence stems from the complex interactions and relationships between ecological systems and human-made systems.

Thus, it is only logical and morally important to care for not only components of the ecosystem, but also their shared relationship with one another. Because individualistic perspectives, even when non-anthropocentric, tend to overlook these systemic relationships, they do not provide a suitable basis for preserving good consequences for ecological systems. At this point, ecocentrism offers a more compelling moral position.

Ecocentric perspectives in ethics place intrinsic value on the well-being of living and non-living natural objects and their relationships (Des Jardins, 2000). An ecocentric position requires that “Wilderness areas, forests, wetlands, prairies, and lakes are valuable in their own right and deserve moral consideration” (Des Jardins, 2000). Such a view is ‘holistic’ in the sense that nature is viewed in its *wholeness*. Instead of prioritising the life forms, an ecocentric approach considers how they interact as a system and how one being’s existence and behaviour influence other beings within the system.

Different ecocentric approaches, such as Aldo Leopold's Land Ethic (1949) and Arne Naess's Deep Ecology (1973), argue for prioritising ecosystems (and thus their components and underlying relationships), as they possess a moral value and good of their own by virtue of their relationships and interdependencies. From this view, an action is right if it contributes to the ecosystemic well-being (its conditions for flourishing and integrity). Thus, an ecocentric position captures the nature of the climate crisis by considering ecosystems as morally valuable[17].

From this view, we have a responsibility to act in ways that uphold the relationships and interdependencies between different ecological systems and their components, as these are all morally valuable. Whilst such arguments are persuasive to some, they have not yet found mainstream philosophical acceptance, let alone outside the realms of academia. Indeed, these ecocentric positions can be criticised for neglecting human happiness or undervaluing human freedom and autonomy. In Chapter 2, a system consequentialist theory that reconciles such human goods with ecological goods will be outlined. Subsequently, this claim about reconciliation will be validated across various parts of this dissertation (for instance, subsection (4.4.4) on principled autonomy and Chapter 6 on trust).

Regardless of their adequacy in capturing the systemic nature of the climate crisis, lines of reasoning–both anthropocentric and non-anthropocentric–require us to recognise a collective moral obligation to address climate change. From both perspectives, we are morally blameworthy if we allow this crisis to worsen and take no action. Anthropocentric reasoning leads to a weaker moral obligation, which may not prompt us to reconsider our relationship with the environment. An obligation resulting from a non-anthropocentric, ecocentric position is stronger and more adequate for our context because it values the

[17] This point is explored in more detail in the next chapter in 2.4.3.

ecosystemic goods and their interrelationships. The nature of such an ecocentric obligation is outlined in the following subsection.

So far, the chapter has shown how climate change is a systemic problem. An anthropocentric position is short-sighted, for it only focuses on humans, ignoring their relationship with the ecosystems. This anthropocentric worldview can be overcome by embracing an ethical approach informed by systems-based thinking, which is essentially an ecocentric ethic in action. As this approach focuses on the value of upholding systemic goodness and addressing climate change requires considering systems as morally valuable, it provides both the correct ethical position and a practical one.

1.5. Human Moral Obligation:

There are uncontroversial moral reasons to act in response to the climate crisis. This subsection focuses on human moral obligation arising from these reasons. First, it shows why such an obligation exists. To have an obligation is to have a duty or a responsibility grounded in moral reasoning. Second, it outlines how this obligation emerges from participation in the systemic harm caused by human actions. Third, it outlines factors that determine the strength of this obligation.

1.5.1. Why do we have a moral obligation?

The destruction of ecosystems and human livelihoods constitutes clear and foreseeable harm; climate injustice distributes that harm unequally, disproportionately affecting those least responsible; and our collective agency as a species renders us accountable for addressing it. The fundamental principles of not harming or causing injustice establish a moral obligation to mitigate and adapt to climate change.

Harm. If we accept, as Adriana Placani does, “we have a prima facie duty not to cause foreseeable and reasonably avoidable harm.” (2025, p. 80), then we have a duty to prevent the harm caused due to climate change. This collective obligation is founded on the

principle of non-maleficence, a relatively uncontroversial principle. Climate change is a human-induced problem. The reports from IPCC and other scientific bodies are unambiguous and highly confident that climate change has already caused some extent of irreversible harm to the Earth's ecosystems[18]. As humans are responsible for this harm, they are already morally blameworthy and have certain remedial moral obligations (Placani, 2025). To say that we have remedial obligations[19] is to acknowledge that we are witnessing a collapse of the ecosystems and are in a position to prevent the conditions from worsening.

The moral imperative to prevent this harm falls on the collective. If we act collectively, we can significantly reduce greenhouse gas emissions and, consequently, substantially mitigate climate-related harms. Thus, our collective obligation exists because we can prevent climate harm through actions such as reframing our economic and political mechanisms, engaging in environmental activism, and implementing large-scale behavioural transitions. Moreover, to the extent that imposing such a collective harm is avoidable[20], we are morally guilty if we impose it (Placani, 2025).

Injustice. Another reason why we have an obligation is that our actions have an unjust impact on certain climate-vulnerable groups. Other living beings, such as flying foxes

[18] For instance, over 50 percent of the Arctic Sea ice has melted since 1979. Certain natural processes, like pollination and carbon sequestration have been compromised in a way that is not reversible in the foreseeable future (IPCC, 2022).

[19] To say we have a remedial obligation is to point out that there is still hope to remedy the environmental damage from developing into a full-scale existential crisis, which can compromise the survival of the natural world and humans. This means that we have to take the responsibility of subduing the damage caused by the past generations. It also means that we leave the scepticism about whether future generations will continue to uphold the moral norms we follow in fulfilling these obligations. Both of these points, regarding taking moral blame for the past and disregarding scepticism about future behaviour, are debatable. Nonetheless, to accept that we have such obligations is the most conducive position to minimise the harm caused by climate change. In prescribing collective obligation or individual obligation, I am in fact, referring to a remedial collective or a remedial individual obligation.

[20] To say that the harm is avoidable means that our collective actions can minimise the risk of such harm. This possibility is unambiguous in scientific reports on climate change. Whether this harm is avoidable on an individual level is a complex issue in relation to individual moral obligations (Nefsky, 2019). It requires looking into the historical, causal, and epistemic aspects of personal habits in relation to climate change. I take this up in the next subsection.

and corals, are already experiencing mass deaths. They have nothing to do with the climate crisis, yet they bear the consequences of human actions.

Moreover, this injustice is also faced by other humans. For instance, to date, the wealthiest quarter of the human population is responsible for three-quarters of emissions that contribute to global temperature rise (Merz et al., 2023; Silva, 2022). The benefits are reaped by the wealthy, whilst the poor bear the costs. More than 3.6 billion people worldwide live in climate-vulnerable regions (IPCC, 2023). The increased effects of climate change will not only lead to direct suffering but will also result in conflicts between vulnerable regions over the acquisition of natural resources (Albert, 2022). These trends are unsettling and signal an uncertain future characterised by instability and suffering for certain groups more than others.

This injustice also affects future generations, who are at risk. The same scientific report, with high confidence, states that the extent of global mean temperatures experienced by future generations depends on the "choices now and in the near term" (IPCC, 2023, p. 6). This question about whether we should act or not, considering the benefits for future generations, gives rise to the intergenerational debate.

Current generations are in a position to exploit the resources meant for future generations. This is to say that our position generates an asymmetric power over the future generations (Gardiner, 2011). This means that the way we act in the present state has implications for people in the future. As our collective actions can cause resource shortages and extreme weather events that impose suffering on non-human beings, vulnerable populations, and future generations, we have a collective moral obligation to act.

Much of the scientific evidence does indicate that human behaviour can mitigate climate change. This means that objective avoidability of the climate crisis is possible, as also accepted by Adriana Placani (2025). As our collective response to address climate change can either improve or worsen the world's situation in the face of climate change, irrespective of

which of the two reasons (harm or injustice) we accept, we have a collective moral obligation.

1.5.2. The Nature of the Obligation:

As a result of this collective obligation, on an individual level, we have a moral responsibility to act in addressing climate change. This view is possible by appealing to a system-based or a putative group-based view.

There are two reasons why individual obligations result from the systemic view. First, one's actions belong to a bigger system that has a significant climate impact. Scholars like Hiller (2011), Robertson (2021), and Placani (2025) appeal to look at individual behaviour as happening within a system. The significance of one's actions is determined by the broader system in which they take place. For instance, even though one's drive does not directly lead to visible climate harms, such as the death of polar bears, one's actions belong to a 'GHG-emitting driving system' that melts the icebergs, leading to habitat loss, which ultimately leads to starvation and death of polar bears (Hiller, 2011). Therefore, one's actions are part of a larger system that causes climate harm and cannot be viewed as separate from the harm inflicted by that system.

Second, one's behaviour underlies a series of behaviours that have implications for the environment. So one's actions inevitably initiate a process of systemic harm. For instance, our impact extends beyond our carbon footprint. Nielsen et al. (2021b) argue that certain individuals, often those belonging to higher socio-economic groups, influence the behaviour of others. These individuals not only act as citizens, but also as decision-makers in organisations, investors, consumers, and role models (Nielsen, Nicholas, et al., 2021b). This suggests that the rippled effects of individual actions can influence others, either by setting an example or by justifying environmentally responsible or harmful behaviour.

Hiller (2011) suggests that when discussing individual obligations, the motivation to change behaviour can be tapped by appealing to the cumulative benefits possible by improving systems, which comprise a loosely connected group[21] of humans having an environmental impact. In a similar sense, Placani (2025) also argues that if we are a part of a putative group that is responsible for larger collective harm, we are morally obligated to take responsibility to cease membership of that group or minimise our impact as much as possible (e.g. by reducing GHG emissions).

To put it simply, if individual acts are causal in driving the systems that contribute to climate change, then one's moral obligation is a derivative of the collective obligations shared by those within these systems. For example, a person driving a gas-guzzling SUV is part of a broader group of individuals who rely on personal transportation, contributing to climate harm. Similarly, when someone drives an expensive SUV, they also serve as an aspirational figure, shaping the behaviour of those around them.

Individual acts of climate harm cumulate into a collective harm, forming informal and often unidentified putative groups whose collective behaviour exerts a systemic environmental impact (Placani, 2025). Thus, individual obligation results from one's participation and impact within a system, characterised as a structured group, such as a corporation, or an unstructured group, like a neighbourhood.

Collectively, we have obligations arising from a duty to maintain the systems. As individuals, we are then obliged to act in ways that will maintain these systems, even if our individual acts seem negligible.[22]

[21] In saying that one violates the moral obligation as a 'group', I mean that each individual is loosely connected as a group. Some common properties of each individual's actions in that group result in a cumulative effect, which imposes significant climate harm.

[22] An extensive literature exists on this issue, highlighting the problems with assigning responsibilities on an individual scale. For instance, the proportionality issue is that the reference to the amount of climate harm is insufficient to make a case for individual obligation (Sinnott-Armstrong, 2003). From the viewpoint of crossing discriminatory thresholds, the fraction of environmental harm imposed on a personal level is too small to generate moral concerns (Garvey, 2011). From a systems-based perspective, one's obligations exist due to the virtue of being complicit in systemically imposed harm.

1.5.3. What does this obligation require us to do?

The causal complexity of climate systems makes it hard to assign responsibility to individual agents, while collective action problems obscure the link between individual action and collective outcomes. For instance, Nefsky (2021) compares the GHG emissions of a gas-guzzling SUV from a person who lives a lavish and carbon-intensive lifestyle versus someone who occasionally takes her Camry out a few times a year for a drive. To say that both actions are equally blameworthy is unfair. A similar view is also held by Nefsky (2021), Placani (2025), Hiller (2011), and Robertson (2021), who hold that an 'imperfect moral obligation' exists on an individual level.

From a systems-based approach, humans (acting within economic, political, and social systems) have obligations that emerge from their impact on the ecological system, which is the foundation of all other functioning systems. We are embedded beings, existing simultaneously as part of multiple systems, both ecological and man-made. However, the intensity, that is, whether the obligation for an individual to act is strong or weak, depends on several factors, some of which are discussed below.

Patterns of Behaviours. The moral wrongness of one's actions can be determined based on how they have acted in the past, not just at an instance or in isolation. This follows an 'imperfect' obligation to avoid acting in "ways that when enough of us act, great harm or injustice results—enough of the time" (Nefsky, 2021, p. 212). Not all actions with an environmental impact are morally blameworthy to the same extent. The moral obligation extends to patterns of behaviour, which, when repeated, significantly contribute to climate change (Placani, 2025; Nefsky, 2021).

Awareness. The extent of awareness about one's role in contributing to climate change also plays a role in the degree to which one violates one's individual moral obligation.

In emphasising the intensity of this obligation, Placani (2025) argues that individuals who knowingly contribute to climate change have, inter alia, a personal moral obligation to minimise their environmental impact.

Humans struggle to have awareness about their moral obligation because climate change is a complex and dynamic process. The temperature shifts driven by climate change are challenging for people to comprehend, calculate, and predict (Kaaronen, 2018). One of the reasons why this has not been possible is because humans do not understand how their behaviour is a causal-driver of climate change, lack the moral will to act sustainably at all times (see also Singer, 1972), and tend to feel less guilty for not acting to address collective action problems (Savulescu & Persson, 2017a; Singer, 1972; Singer, 2008). Chapter 3 considers the psychological factors that underlie this limitation of human behaviour in practice, and examines whether nudging can provide a solution.

Improving behavioural outcomes, patterns, and awareness is necessary to mitigate climate change in the long term. This can be done at a collective level (i.e., at the level of states, markets, etc.) or at the individual level. Obviously enough, individual changes will play a role in leading to collective behavioural changes.

Individuals acting on their moral obligation will involve reducing GHG emissions through direct changes, such as driving less, flying less, or reducing their meat consumption. However, simply acting is not enough, considering the state of climate change (Nielsen et al., 2021a; Nielsen et al., 2024). Along with direct changes, indirect influences, such as our socio-economic privilege, can motivate others to reduce their climate impact (Nielsen et al., 2021b). In practice, driving behavioural shifts, developing moral motivation, and awareness about one's behaviour have proven to be a Herculean task for governments, international organisations, and groups (Nielsen, Nicholas, et al., 2021a; Lehner et al., 2016).

As such, it appears that there exists an obligation to consider "Green Nudging" (GN), i.e. nudging designed to promote environmentally beneficial behaviours. GN can help us prevent harm and achieve better systemic outcomes for both humanity and non-human nature. Ergo, we have an obligation to pursue it. GN is a strategy that can not only drive behavioural change towards climate-friendly solutions but also lead to increased awareness about our relationship with the environment, and subsequently help us fulfil our moral obligation in addressing the issue.

1.6. Green Nudging as a Solution:

Nudging has been developed as a distinct strategy to improve behavioural outcomes– informed by behavioural sciences and psychology. Nudging was popularised by Thaler and Sunstein (2008) in their book, *Nudge: Improving Decisions About Health, Wealth, and Happiness.* They define a nudge as an intervention that "alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, 2008, p. 6). To nudge is to influence one's choices towards better decisions through stimuli or changes in the decision-making context (Thaler & Sunstein, 2008; Savulescu & Persson, 2012). The precise definition of what constitutes a nudge has been a point of debate amongst scholars working in this field. Simply put, to nudge someone is to change the way they perceive their choices at a given time, by highlighting specific options, providing them with information, appealing to their moral values, or appealing to social norms. GN, then, in the same way as nudges, improves behaviour to accomplish pro-environmental outcomes. Chapter 3 discusses the scope of nudging and GN in relation to this thesis.

Considering the urgency of improving our response to climate change, influencing human behaviour is a pragmatic starting point. A focus on individual behaviour change is feasible, timely, and useful. GN can potentially act as systemic interventions that shift

collective behavioural patterns, making our human systems functionally efficient in upholding ecological well-being. Thus far, philosophical and religious education has only had limited success in making us "more environmental", but GN may offer the possibility of helping align individual and collective behaviour with the maintenance of ecosystemic goods in three ways: behavioural change, awareness, and cognitive shift.

First, GN directs individuals toward environmentally friendly behaviour by influencing areas like mobility, energy, and food habits (Lenzi et al., 2016). For example, personalised energy feedback for mobility reduced GHG emissions by 19-35% in Japan (Fuji & Taniguchi, 2006). For further context, in this study, participants completed surveys and travel diaries, consenting to be nudged through strategies such as personal feedback and social norms (Fuji & Taniguchi, 2006).

Similarly, reducing plate size in a Nordic hotel chain resulted in a 15-20% reduction in food waste (Kallbekken & Sælen, 2013). Moreover, specific categories of GN are directed towards reducing meat consumption, because the procurement and processing of meat contribute to climate change. Within controlled environments, such as canteens and restaurants, nudges that alter the visibility of meat-based options, communicate social norms, and simplify information have shown promise in influencing food consumption behaviour (Lehner et al., 2016). These studies serve as examples of how GN is effective in changing behaviours for climate change mitigation (in Chapter 3).

Second, GN can help individuals become aware of how their actions contribute to a series of actions that cause climate harm. For instance, leveraging social norms based on user data, incorporating reflection along with the use of defaults or personalised nudging are some of the strategies that have sought to reform the communication of the climate change agenda through nudging and behavioural insights (Banerjee & John, 2021; DellaValle & Sareen, 2020). Framing information in a manner that informs an individual of personal benefits and

reducing emissions due to an action informs environmental decision-making in individuals being nudged (Cappa et al., 2020).

Similarly, nudging can be a powerful way to visualise the effect of one's actions on the behaviour of others. With the advancement of intelligent technology, GN can be personalised through the Internet of Things (IoT) (Ranchordas, 2019). This means that aspects of one's daily life, such as driving, travelling, food consumption, and shopping, can be tailored through lifestyle applications without incurring monetary costs, ultimately minimising one's negative impact on the environment and raising awareness about their environmental footprint.

Third, on a psychological level, nudging can reconfigure how humans perceive their contribution to climate change, ultimately helping us win the 'cognitive war' with this collective crisis (Kaaronen, 2018). Kaaronen (2018) uses this term to refer to the psychological struggle of humans to come to terms with the state of climate change and its transition into a global crisis. Nudging can help us shift the way we perceive our relationship with the environment, and thus reframe our understanding of the moral responsibilities in addressing climate change (Kaaronen, 2018; Hukkinen, 2016).

For instance, Kaaronen's (2018) example illustrates how this reconfiguration is possible. A daily activity, such as buying groceries from a supermarket, is perceived by humans as the goal of buying certain items. A reconfigured cognitive system would perceive shopping not as a single activity, but rather as a series of interconnected events that have a systemic impact. Shopping is viewed as a process where one chooses the means of transportation to use, the products to purchase, and the environmental impact of such consumption and disposal, among other factors. This shift is possible through nudging strategies that can address two psychological limitations that obscure our awareness of climate-related harms due to our choices. This proposal of a shift in human psychology and

subsequent changes in behaviour form part of this examination that explores the ethics of GN.

1.7. Conclusion:

This chapter explored how climate change is a behavioural issue. It argued that the failure to respond to the climate crisis appropriately is a lapse in following both collective and individual moral obligations. In what follows, the next chapter (Chapter 2) formulates the theory of *Environmental System Consequentialism*, which is essentially a 'rule-based objective list system consequentialism'. This is the primary theoretical apparatus used in this thesis to frame the climate crisis and propose an ethical framework for GN.

Chapter 3 examines how GN can be effective in addressing psychological factors that impede effective climate action. Despite the potential, there are ethical concerns associated with the use of GN, which will be outlined in Chapter 4. This chapter will also explore whether the severity of the climate crisis justifies overlooking some of these ethical issues.

With these two arguments–individual obligation to address climate change and the effectiveness of GN–I can offer a system consequentialist defence (Hiller 2014, 2016) of GN in Chapter 5. This chapter argues that an ethical justification of GN does not lie only in isolated behavioural changes, but in how these behavioural shifts improve systemic outcomes in mitigating climate change. This systemic justification is driven by the potential of GN to (1) change behaviours, (2) improve awareness about the ethical implications of one's behaviour, and (3) drive the cognitive shift to enhance human relationships with the natural world.

The thesis then examines the system of nudging between two parties, arguing that trust between the person being nudged (the nudgee) and the agency using nudges (the nudger) generates the conditions required for ethically justifiable GN in Chapter 6. Such a proposal

would be of interest to pragmatists, environmental philosophers, and policymakers alike who strive to improve pro-environmental behaviour. Lastly, Chapter 7 provides procedural recommendations for designing effective GN and illustrates communication strategies that can embed GN into a self-governance framework.

Chapter 2. Environmental System Consequentialism

As illustrated in Chapter 1 (1.3), the effects of climate change are systemic stressors to the Earth system. Different subsystems, e.g., ecosystems, atmospheric systems, economic systems, and political systems, that belong to and are interconnected in the Earth system experience the consequences of human actions. A fact, then, is apparent: human activities undermine the well-being of all these systems.

Human contributions to climate change lead to bad consequences for the Earth system overall, marked by ecological instability, economic and social stressors, and biodiversity loss. Arguably, the nature of the problem prescribes a consequentialist approach. As GN may help mitigate this threat to the functioning of ecological systems, which directly bear the effects of human actions, we need to consider whether GN constitutes an ethical strategy for navigating this global challenge.

In this chapter, I argue for *Environmental System Consequentialism* (ESC), the theoretical apparatus for looking at the climate crisis, which is then applied to develop an ethical framework for GN (in Chapter 5). To develop ESC in this chapter, the reasoning for consequentialism and systems-based thinking for climate change is explored (2.1), the methods of assessment (in the consequentialist schools of thought) (Section 2.2), and measures of consequentialist values are also outlined (Section 2.3). Then, a pluralistic consequentialist approach, ESC, is formulated (2.4). It is shown how ESC provides a basis for considering ecosystems as morally valuable, providing a coherent consequentialist position for examining the role of human actions in contributing to climate change.

2.1. Consequentialism and Climate Change:

Climate change is a moral issue because it involves practical concerns and normative concerns (discussed in Chapter 1): it raises the questions of right and wrong about what

actions can prevent bad consequences for humans and ecological systems together in the present as well as future. As climate change involves immediate and long-term harms to complex and interdependent systems (as well as beings who are dependent on those systems), an outcome-focused consequentialist position about systems is well-suited for examining the climate crisis.

For consequentialists, the morality of an act is based on whether the consequences (or the state of affairs) resulting from that action are morally good or bad. Simply put, consequentialism commits us to the position that an action is morally right based on whether the consequences resulting from that action are good. "Consequentialism, as its name suggests, is simply the view that normative properties depend only on consequences" (Sinnott-Armstrong, 2023) and,

> …the most prominent example is probably consequentialism about the moral rightness of acts, which holds that whether an act is morally right depends only on the consequences of that act or of something related to that act, such as the motive behind the act or a general rule requiring acts of the same kind. (Sinnott-Armstrong, 2023, para. 01).

In other words, for consequentialists, an action is morally justified if it leads to the best possible outcomes or state of affairs.

Other traditional moral theories that focus solely on character or duties struggle to capture the full complexity of climate change as a systemic problem. For instance, there are three distinct reasons why a deontological framework such as Paul Taylor's (1981) faces limitations. First, by ascribing equal moral standing to all organisms, an egalitarian rights-based approach is likely to result in irresolvable rights conflicts. As climate change is a deeply intertwined and systemic problem, trade-offs are unavoidable. Second, trading-off is further complicated by the fact that deontological frameworks consider individual rights and

duties as a primary unit of moral concern, whereas climate change is a systemic issue. When moral accountability is distributed across collective, cumulative, and generational actions rather than individual agents, deontological frameworks can lead to accountability deficits. Third, merely focusing on rules can impede effective response due to dynamic and slow-moving character of climate change. For these reasons, even though deontological insights on autonomy are valuable to ESC (as we will see in Chapter 5), a consequentialist framework seems better suited for navigating the systemic nature of climate change.

Similarly, given the urgency of climate change, relying solely on character development by adopting a virtue-based position is impractical for two reasons. First, virtue ethics poses a traceability issue: when GN operate at population-levels, it is difficult to assess whether GN truly foster virtue or not. Second, a virtue-based position is better suited to evaluating long-term character development rather than immediate outcomes, yet the urgency of climate change requires immediate measures as well.

A consequentialist position can look at GN based on their immediate outcomes and their ability to foster virtues that are instrumental for good consequences. This is not to say that virtue ethics cannot provide a systemic defence of GN. Nonetheless, an ESC position is preferable because it is more pragmatic and direct than a deontological or virtue-ethical position in considering systemic outcomes. The climate crisis is urgent and arises from the interplay of ecological and human systems and therefore demands a moral perspective capable of analysing systems as such.

For such reasons, a consequentialist framework, *Environmental System Consequentialism,* is adopted in this thesis. Without dismissal of the deontological or virtue-based positions (which can be compatible with ESC to some extent), the systemic and outcome-focused nature of the climate crisis calls for a consequentialist framework. However, it is worth acknowledging that a consequentialist position has its own drawbacks

regarding measurability of systemic outcomes and assumptions about human nature and collective agency, which this thesis acknowledges, works within, and yet does not fully resolve. Chapter 3 explores whether a change in human nature and subsequently in collective agency is possible through GN, but the proposal remains pragmatic rather than definitive.

Having argued for a consequentialist framework, it is also necessary to consider the relationship between ESC and the ecocentric line of thought. Ecocentric approaches, such as in the works of philosophers like Arne Naess (2005) and Holmes Rolston III (2011), prescribe moral judgements on the basis of how they contribute to the ecosystems. ESC adapts the line of thinking reflected in ecocentric thought–the idea that ecosystems hold intrinsic value–while grounding ESC in a consequentialist framework.

Nonetheless, despite emerging from an ecocentric approach, ESC is different from traditional ecocentric approaches for three reasons. First, Naess's appeal for deeper transformation through 'oneness' or Rolston's focus on holism do not provide a normative guide about how individual and collective actions should be directed to mitigate the climate crisis. This poses a limitation when the goal is to assess whether an individually-focused strategy like GN can be ethically justified to address the collective crisis of climate change. Second, by fully rejecting anthropocentrism, these approaches do not adequately frame human moral accountability[23] in the face of the climate crisis, a problem that is anthropogenic in nature and requires human intervention to be resolved. Third, ecocentric theories, although they value ecosystems, are so radical in their scope and conditions that, arguably, they are unlikely to be politically or socially acceptable at any time in the near future. For instance, these positions may subscribe to some radical ways to address the problem of climate change, like reducing the human population by killing, coercing humans not to procreate, or shutting

[23] For instance, these ecocentric approaches do not prescribe what we ought to act in our everyday life to tackle a problem that is so systemic and deeply embedded in our daily lives.

down all activities that have an environmental impact (and leaving a great number of people in dire social and economic circumstances) (Hiller, 2013)

In looking at climate change, we need a moral theory grounded in systems thinking — one that evaluates behaviours, attitudes, and obligations in terms of the extent to which actions contribute to good outcomes at the ecosystemic level. ESC therefore synthesises ecocentric moral concern for ecosystems in a consequentialist framework that can guide human action, preserving what ecocentrism gets right while providing the normative framework it lacks.

In other words, in considering moral dilemmas, we should try to think about which ecosystemic and other human-centric systemic goods are at stake and how they can be preserved. The aim in this chapter is to achieve a balance between ecocentrism and anthropocentric values, and to provide a synthesis that will serve as the pragmatic basis for human action in response to climate change, whilst also focusing on the intrinsic value of ecosystems.

This chapter will argue for a pluralistic consequentialist theory of ecological systems that provides a coherent and rational basis for acting ethically. This is to say that the version of system consequentialism formulated in this chapter maximises the good outcomes for the ecological systems, but at the same time upholds what is morally good for the other valuable systems for reasons discussed in 2.4.2. In short, the version of *ESC* developed in this chapter is both practical and plausible for responding to the climate crisis.

For the sake of this dissertation, this is important because it provides a set of principles to examine whether GN is ethically justified. When considering GN as a response to climate change, it is essential that we consider both what is beneficial for ecological systems, whilst also considering the functioning and stability of human systems that subsist on these. This means that a defence of GN from this position will establish GN as a tool that

maximise the good outcomes for the ecological systems, while at the same time providing conditions that preserve other systems necessary for human well-being (this theory is applied to GN in Chapter 5).

2.2. Indirect and Direct Consequentialism:

There are two ways of evaluating consequences: an indirect or a direct form of consequentialism. A direct consequentialist, or a *global direct consequentialist*, holds that the moral quality of something (an action, a motive, or a virtue) depends on its consequences, as opposed to the consequences of everything else (the agent's motives, rules, or virtues) (Sinott-Armstrong, 2003). A direct consequentialist about motives will find a motive to be morally justified if it leads to the best consequences. The most common form of direct consequentialism is Jeremy Bentham's act utilitarianism*,* which holds that an act is morally justified depending on whether it maximises pleasure or minimises pain, which are consequentialist values for a utilitarian position.

An indirect consequentialist holds that the moral justification of something (a motive, virtue, or rule) depends on the consequences of something else (Sinott-Armstrong, 2003). For example, an indirect consequentialist about motives would argue that an action is morally right if it stems from a motive that generates the best possible outcomes. A prominent form of indirect consequentialism is rule consequentialism, which evaluates actions based on the consequences of the rules that permit them (although John Stuart Mill is often associated with ideas resembling *rule utilitarianism*) (Kuusela, 2011; Sinott-Armstrong, 2003). Rule utilitarianism holds that an act is morally right if it results from a rule that leads to the best state of affairs, by maximising pleasure or minimising pain, or satisfying a person's preferences insofar as they are compatible with others' preferences being satisfied.

Importantly, this means that an act is permissible when it follows a *rule* that maximises good rather than focusing on the actual outcomes of the act.

Most consequentialists maintain that it is better to follow general rules that have been established to maximise the good, rather than assessing every action in terms of direct outcomes (Sinnott-Armstrong, 2023). These rules might focus on maximising good motives, virtues, welfare, preference satisfaction, or a system to achieve good consequences. This is known as indirect consequentialism.

Rules can work as guides in law, governance, or administration to help us maximise the good outcomes. An indirect form of consequentialism is generalisable (Sinnott-Armstrong, 2023). In other words, an indirect assessment of consequences works better because there are practical limitations when it comes to acting without rules. For instance, it is better to follow specific rules, such as "Do not lie.", "Do not kill.", or "Do not cheat.", instead of evaluating the consequences of doing so every time.

Moreover, through indirect consequentialism, rules are useful because they can achieve values that maximise long-term good consequences for the ecosystems[24]. Certain objective goods, like ecological health, balance, and relationships, are necessary for upholding ecosystemic well-being while preserving other vital systems (legal, social, and political systems). Societal virtues like trust, justice, and cooperation can be preserved if we follow rules about them and others do the same. Thus, an indirect consequentialist position is preferable as it can focus on achieving consistent good outcomes through rules.

[24]One way to look at this is that with limited knowledge in a circumstance, there are constraints on our ability to maximise the good outcomes, both long-term and immediate. This is why a direct consequentialist position falls short in this context.

Consider an example to illustrate this. We know that medical practitioners act according to certain rules that are thought to maximise good consequences. For instance, most of us trust that our doctor would not transplant one of our kidneys to a person who is extremely sick and about to die. We generally believe that doctors will not break certain rules regarding patient welfare and consent, even if acting in such ways does not always maximise good consequences. Following rules upholds values like trust and cooperation among patients within the medical system. In the absence of such rules, society would collapse. No one would trust the medical system. There are many examples of how some values are pursued indirectly when one acts according to rules about honesty, fidelity, or privacy.

There are two reasons why considering the consequences of rules of action, rather than direct consequences, gives us a stronger moral position and should be our focus for examining the actions in relation to climate change, and therefore the ethics of GN.

First, rule consequentialism will prescribe consistent actions to mitigate climate change. This may involve focusing on generating pro-environmental behavioural outcomes in practice through certain habits, patterns, or attitudes. A direct consequentialist position will not provide policymakers with general thumb rules for maximising good consequences. Instead, they will have to merit each instance specifically to generate the best possible outcomes through each environmental intervention. If they are unable to do so, their actions to mitigate climate change cannot be justified through direct consequentialist positions, even if they are aimed at achieving good consequences. This is practically difficult.

An indirect consequentialist justification of any climate policy would find the consequences to be morally justified based on whether it results from policymakers or decision-makers acting according to rules about something that maximises good consequences. For instance, Rawls (1955) argues that policymakers should act according to the rules that are built into institutions and are defensible in public. This is feasible in

practice, especially for GN, as compared to the alternative of direct consequentialist justification. To embed GN into governance practice and to cover them under legal mechanisms, we must establish rules regarding good practices for GN.

Second, a direct consequentialist may prescribe that we ought to fulfil moral obligations to generate the best possible outcomes for the ecosystemic good. However, the issue with direct consequentialism is that it does not explain why preserving certain values that are important for other vital systems is essential, especially if the alternatives can generate equally good outcomes. For instance, if there is an option of using a more manipulative nudging strategy that is less transparent but achieves the same result, a direct consequentialist position will not tell us why we should prefer the one that upholds values like autonomy, trust, or fairness, which are required for other systems, like a democratic one.

An indirect position allows for a holistic approach consistent with ecocentric perspectives. Establishing rules about good outcomes helps protect systems that both affect and depend on ecological health. Specific objective values, as we will see, can help us make rules in this regard. This framework has implications not only for individual obligations towards the ecosystem but also for how people act within other social and economic systems, which are vital. For instance, from a system consequentialist perspective, it would be desirable that, as democratic agents, people are more aware of their climate impact. As moral agents, they better understand their relationship with the environment and have more environmentally conscious preferences. Thus, the best state of affairs associated with decisions concerning the environment and thus GN (as we will see in Chapter 5) should be assessed from an indirect consequentialist position. The next subsection considers different consequentialist values.

2.3. Consequentialist Values:

Another question is, “What constitutes the best state of affairs?”. Various existing consequentialist theories use different measures of value. That is, they consider different measures of values for determining whether an action is right. A theory can prescribe one or two supreme consequentialist values, or a range of values. For instance, pleasure and pain are considered morally valuable, where both are correlative, simple, and unitary concepts according to the Hedonists (Maclyntyre, 1966). On the other hand, a range of properties can be considered morally valuable, such as the Objective list of values proposed by Derek Parfit (1984). In what follows, different consequentialist values are explored, and a notion of ‘systemic values’ is claimed, which is similar to the objective list consequentialist approach.

2.3.1. Hedonism

For Hedonists, an action that brings about the best state of affairs is the one that maximises pleasure or minimises pain. In other words, for hedonists, pleasure and pain are the main consequentialist values. As pain is the absence of pleasure, both are correlative and unitary concepts (Scruton, 2002; MacIntyre, 1966).

While the history of hedonism can be traced back to proto-utilitarians such as Hobbes and Hume, Jeremy Bentham is widely regarded as the primary architect of act utilitarianism, a hedonistic view that regards pleasure as the ultimate good (Scruton, 2002). Bentham notes that we have been placed under the “governance of two sovereign masters, pain and pleasure.” (1789, Chapter 1, para. I). In his view, the pursuit of the best state of affairs cannot be independent of the experience of pleasure and pain.

Maximising the morally good consequences, according to Bentham, is possible by summing the pleasure experienced by each individual in the society (Bentham, Chapter 1, para.V). Individuals make up the society, where each one is in pursuit of their own good or pleasure, and we can calculate the total good in the society by summing the happiness of each individual (MacIntyre, 1966; Goodin, 1993). Thus, for utilitarianism, according to Bentham,

the greatest good lies in acting in ways that seek the greatest amount of happiness for the greatest number of individuals (Bentham, Chapter 1;MacIntyre, 1966).

An issue with Bentham's utilitarianism is the quantification of pleasure and pain (Copp, 2006). Although Bentham identifies properties of pleasures that are morally desirable, including pleasures of the senses, pleasures of skill, and pleasures of benevolence (Bentham, Chapter V, para. II), his account does not assess the quality of pleasure. For instance, it does not tell us whether greater happiness results from solving a crossword puzzle or saving a drowning person.

John Stuart Mill, a hedonist who was the architect of rule utilitarianism, introduces a qualitative distinction between higher and lower pleasures. Mill argues for rule utilitarianism (a method of assessing the consequences) discussed in the previous subsection, according to which we must act according to the rules to maximise pleasure. For Mill, pleasure is a unitary concept for judging the moral rightness of an action. However, he holds that not all kinds of pleasures are equally valuable (Mill, 2006). He famously writes, "It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied." (Mill, 2006, p. 70). Mill holds a hedonist position which posits that certain kinds of pleasures are qualitatively different from other pleasures. According to his account, in seeking the greatest happiness for the greatest number of individuals, we must also consider the quality of pleasure being derived from one's actions.

There are two main problems with hedonism. First, the question of whether pleasures and pains are psychological states, and if they are felt by everyone in the same manner. If we recognise the fact that human sensations of pleasure and pain are framed by one's conditioning or psychological dispositions, then this concept of good is morally dangerous (Copp, 2006; MacIntyre, 1966). Human nature is malleable, and we can be conditioned to

accept pleasure and pain in situations that are dangerous[25]. For instance, torturers derive pleasure from killing others for sport or enjoying violence. This view has ramifications in the bigger picture: hedonism can justify totalitarianism, where individuals can be trained into ignoring the lack of values like freedom, peace, and democratic participation (MacIntyre, 1966). Moreover, climate change is an example where conspicuous consumption of goods provides most of us with what we identify as pleasure, and is seen as contributing to a good life, even if that has a negative environmental impact.

Second, hedonism is a desire-based consequentialist because associates good outcomes with a moral agent's appetites (Dietz, 2019). A desire-based consequentialist theory associates good outcomes based on people getting what they desire (Copp, 2006). In other words, it evaluates actions by how well they satisfy people's desires, rather than by pleasure, preferences, or other goods. This is a problem because for these positions, desires play the central role in the nature and the possibility of pleasure (or good outcomes); this fact explains this problem (Dietz, 2019). In the case of the climate crisis and the human experience of this issue, having the right appetite or desires to rationally assess the possibility of good outcomes is not possible. As outlined in (1.3.1) and will be further developed in the next chapter, human psychology, trained by evolutionary factors, is inept to deal with climate change (Steg, 2023). Thus in evaluating the moral rightness of an action concerning the ecosystem, taking hedonist perceptions of pleasure and pain will not suffice for ensuring the best possible outcomes.

This suggests that a type of 'objectivity' in consequences is needed–an experience that is true, and not only a psychological sensation. In summary, for hedonists, moral goods can be acquired through the psychological dispositions that lead to experiencing pleasure or

[25] This will be a theme in Chapter 3, where it will be shown how certain psychological traits make us apathetic towards the harm and suffering faced by others. GN can help us deal with those.

pain; however, this overlooks the broader question of whether sensations of pleasure and pain are the right means of determining the best state of affairs in light of the climate crisis.

2.3.2. Preference Satisfaction:

According to a preference-based account of consequentialism, the ultimate measure of value lies in the satisfaction of individuals' preferences or desires, irrespective of whether those preferences lead to pleasure or avoid pain. This is also a desire-based consequentialist position. While people may prefer pleasure and pain to satisfy their desires, preference consequentialists argue that what is valuable is whether a person's considered preferences are fulfilled. For instance, one may derive pleasure from certain unhealthy foods, but they still prefer to avoid them for health-related reasons. So what makes an outcome good is not the psychological sensation of getting pleasure from eating, but their preference for living a healthier life.

In considering Impartiality[26], which is a critical consideration for a good ethical theory, Singer (1993) suggests that all critical judgements should be made from a point of view where my own interests count no more than the interests of anyone else. Everybody has an equal right to have their preferences satisfied. If their preferences are at odds, then one should weigh them equally, taking into account the preferences of those who are affected by the decision (Singer, 1993). One must choose the course of action that best satisfies the preferences of all those affected by an action.

Peter Singer (1993) and R.M. Hare (1994) argue for the satisfaction or dissatisfaction of preferences as the basis of moral judgements. This condition addresses a critical objection to hedonism: the assumption that what gives me pleasure is naturally pleasurable to others.

[26] Bentham's and Mill's versions of consequentialism are also impartial positions, holding that 'Everyone counts for one, no one to count more than one'.

The sensation of pleasure and pain can be felt differently by individuals. Thus, the experience of pleasure and pain is not universalisable (Copp, 2006; Goodin, 1993).

Peter Singer (1993) argues that individuals do not have 'utility meters' attached to them that can objectively justify pleasure and pain on a measured basis (Goodin, 1993). This version of consequentialism recognises that different people derive pleasure or moral goodness from different things. For instance, one can find playing outside in the rain is pleasurable, but for another, it might not be a joyful activity. The psychological states of pleasure and pain will be expressed in terms of preferences. Considering the satisfaction of preferences as a moral good allows this theory to have a universally applicable measure of consequentialist value.

Preference consequentialists prescribe satisfaction of preferences, and not just our own, but all beings who are affected by the outcomes of an action. They argue that all preferences ought to be considered equally, so this position is impartial. However, not all beings are capable of having preferences, or at least we have been unable to find scientific evidence for their sentience[27]. So from this position, many beings depend on ecological systems for survival, and harming ecosystems compromises their preferences (e.g., the desire not to suffer). The problem with considering preferences as the ultimate measure of value is that, in the absence of expressed, reasonable, or recognised preferences, this consequentialist position does not take us far in achieving a fundamental good or the best state of affairs.

This is especially evident in the context of climate change. As humans, we can only anticipate what others, who are subjected to the effects of climate change, prefer. Moreover,

[27] Note that the categorisation of certain beings as 'sentient' has epistemological limitations. This means that our current scientific methods and human capabilities do not yet provide a robust basis to ascertain whether other living beings can experience pleasure or pain. So what we consider to be sentient based on scientific knowledge is open to falsification. This issue lies at the core of empirical observation and scientific methodologies. As such, this remains an important area of research in philosophy of science.

psychology has shown, and as we will explore in the next chapter, human moral psychology is shaped by evolution. Humans show tendencies for being selfish, short-sighted, and limited in their understanding of how our actions have implications for the preferences of others (Persson & Savulescu, 2015). If given the option, humans tend to be biased, focusing on their own well-being at the expense of others. This theme is also recurring in Singer's (1981) work, where he argues that we need to expand our moral circle and consider the preferences of others as important as our own. Singer's appeal is morally desirable[28]. However, practically, this is difficult for humans.

Both hedonism and preference satisfaction are desire-informed consequentialist theories (Copp, 2006). A type of 'psychologism' associated with these positions, that is, the psychological experience of pleasure or preference, limits their potential in addressing the problem of climate change adequately. This is because the bad outcomes resulting from climate change surpass human psychological capabilities to perceive those consequences. Another type of consequentialism involves perfectionist theories, such as Derek Parfit's Objective List (1984), which claim that morally valuable consequences can be achieved by pursuing some form of objective human good (Copp, 2006).

2.3.3. Objective List Consequentialism

Certain intrinsic values are required for the "best state of affairs" to exist, irrespective of our desire for them. This is a perfectionist view (Copp, 2007). For instance, for someone to live a truly good life, they must have good friendships, basic amenities of food and shelter, a sense of agency, and a sense of accomplishment in life, irrespective of whether they explicitly have a preference for them (Mulgan, 2014). To live a good life, one must also avoid certain bad states, like excessive intoxication, hatred, or violence. Without such

[28] This is to say that if there were a scientifically robust basis to communicate and calculate the preferences of all beings, and even the preferences of the ecological systems as a whole, preference satisfaction would provide a workable basis as a consequentialist good. But this is not the case, so we consider some objective consequences that can be seen as the basis of a universal moral good.

intrinsic goods, a "best state of affairs" is not possible. This means that certain objective states, like friendships, love, nutrition, and dignity, are required to ensure the best state of affairs. This view is held by objective value list consequentialists who argue that the experience of pleasure and pain is not sufficient as the ultimate good due to limitations in one's perception, recognition, or understanding of these values.

This points out what is essentially wrong with desire-informed consequentialist theories. Imagine if I meet an individual who is poor and suffering, and I express my preference for them to overcome their state of poverty, but I do nothing about it. If I never see this person again and their life gets better without me ever finding out about it, does that make my own life better? Parfit (1984) argues that this satisfaction of my desires does not make a difference to the quality of my life. A type of subjectivism associated with the fulfilment of desires or feelings of pleasure and pain ignores the idea that a morally good life can exist beyond experiences of these sensations (Parfit, 1984).

This position is particularly relevant in the context of climate change, where certain intrinsic goods related to flourishing and well-being are essential for the goodness of all systems. These conditions may not be observable by humans properly[29]; nonetheless, they exist. As shown in Chapter 1, the climate crisis, driven by human actions, risks these intrinsic goods. As we have a moral obligation to address this issue on both a collective and an individual level, we also have a moral obligation to protect these objective, intrinsic goods for the ecosystems.

In short, objective consequentialists, such as Derek Parfit (1984), James Griffin (1986), and Tim Mulgan (2014), argue that several values are necessary to achieve consequentialist goods. A lot of these values may not be desired by a person, but they are

[29] This idea is developed in the next chapter. Two psychological hurdles, *Epistemic Mypoia* and *Diffused sense of Moral obligation,* prevent us from understanding the value of actions that are morally right concerning a collective crisis like climate change.

valuable to accomplish happiness and welfare. In what follows, I claim a similar consequentialist position that demands values that are good for a system.

2.4. Developing the theory of Environmental System Consequentialism:

The way we experience pleasure, pain, or preferences (both our own and those of others) is ill-suited to address the collective crisis of climate change. As briefly outlined in Chapter 1 (1.3.1) and will be explored in the next chapter (Chapter 3), there are evolutionary reasons behind this.

Despite being aware of the facts of climate change, as suggested by scientifically robust reports as well as our own experience of changing weather patterns due to climate change, we are unable to derive pleasure or preference from acting to save millions who are at risk due to climate change.[30] This disconnect diminishes our sense of obligation to act to reduce our climate impact. Thus, climate change exposes the limitations of desire-based theories: these theories rely on preferences or desires that we may not yet possess to deal with climate change.

In the case of climate change, preference consequentialists may still argue that we can change preferences radically to help one recognise the impact of their actions. There is, however, a caveat. To make a moral appeal to change preferences, they must first subscribe to the position that there are certain objective goods (or objective desires/preferences) that are essential for maximising good outcomes. That is, preference consequentialists must commit to improving values such as empathy, sense of justice, and altruism in people (to effectively address climate change). In other words, preference consequentialists must endorse certain objective goods that can improve human capacities to possess well-informed preferences or desires.

[30] GN can help us act despite these psychological hurdles. This point is also applicable in case of carbon taxes (See: (Mintz-Woo, 2025))

Thus, such a position would require them to approve certain objective goods that are necessary for achieving the best state of affairs for the ecological systems. Simply put, a preference-based consequentialist (or a hedonist) will have to subscribe to an objective list to argue for improvements of one's preferences or desires that makes them more capable to understand their contribution to climate change (which can maximise long-term ecosystemic outcomes). This discussion is continued in subsection (2.4.2) in light of the system consequentialist position.

Thus far, we have seen that climate change has implications for ecosystems, which are morally valuable systems on which life depends. Ecosystems are also pivotal for other systems to exist. In addressing climate change, then, we ought to act in ways that preserve the good outcomes for the ecosystem. These outcomes can be considered through a consequentialist theory focusing on systemic outcomes.

System consequentialism holds that an action is morally justified based on its contribution to the overall goodness of a system. This means that for system consequentialists, systemic well-being is the measure of value (Hiller, 2011). By adopting systemic goodness as the consequentialist measure of value, we are positioned to view climate change as a problem that challenges the well-being, flourishing, and integrity of ecosystems. In other words, this position treats ecosystemic health as morally valuable consequences of action.

Next, three subsections formulate the theory of *Environmental System Consequentialism*: (2.4.1) outlines the main tenets of System Consequentialism and its background, as formulated by Avram Hiller (2013, 2016), (2.4.2) shows how system consequentialism combines with objective list, (2.4.3) argues for ecosystemic goods as consequentialist values needed to maximise good outcomes for the ecosystems, (2.3.4)

presents a rules-based objective list version of system consequentialism theory, called *Environmental System Consequentialism* in this thesis.

2.4.1. Tenets of System Consequentialism

The general maxim from System Consequentialism holds that: *If we have an option to do x that leads to better systemic consequences, we must prefer x over other available options.* This formal argument implies that between two actions, where one contributes to the well-being of the system more than the other, the former has moral justification according to system consequentialist principles. This theory can be applied to almost all activities by viewing them as part of a system, including competitions, personal finances, the environment, and health. For instance, if there is a system of competition, fair play respects the integrity of that system, rather than seeking dishonest alternatives that may lead to victory.

A system, in this context, is a 'first-order theorising' attempt as a focus on the systems (in climate change context) does not tap into metaphysical aspects[31], such as where to define the boundaries of the system, who gets to be a part of the system, and what kind of relationships justify interconnectedness in a system (Hiller, 2016). Hiller applies systems thinking to determine whether an action is morally justified. He defines a system as a "whole which has, as parts, interacting individuals which are tokens of types which have a necessary connection to each other" (Hiller, 2013, p. 91).

Consider an example to illustrate this. A forest is a *type* of forest ecosystem. The forest has different *tokens*, which include deer, trees, soil, rocks, lions, and colourful birds. Each of these *tokens* is a *type*. For instance, a deer in this ecosystem is a *token* of the *type* deer. A lion in this forest belongs to the *type,* lion, and so on. These *types* have specific roles: Trees and grass grow on the soil. Animals like deer feed on the grass. Bigger predators, such

[31] In the context of this dissertation, scientific evidence is unambiguous that climate change is a systemic issue.

as lions, feed on deer. Within the forest ecosystem, these *types* (grass, deer, lions) have a definite role to play, and they are interconnected with each other.

To understand and make decisions about a system, two epistemic conditions are necessary. The first condition is that the various *tokens* and their relationship in a system are not accidental. For a moral agent to make a decision that is consistent with system consequentialism, one requires knowledge about the relationship between tokens of a system. If the relationships are accidental, then one would not know what actions are a means to achieve systemic goodness.

The second condition is that the *tokens* of the system must belong to a *type* known to the moral agent (Hiller, 2013). So, in a system, reference to *tokens* must be meaningful. For example, a *token,* say a deer, belongs to a *type* called deer, or more specifically, to a *type* of a herbivorous animal. Specifying the *types* of *tokens* in a system allows a moral agent to draw relationships between different components of an ecosystem, and predict actions that maximise systemic goodness.

These two conditions are fulfilled in relation to human behaviour and climate change. First, we have sufficient scientific evidence that our actions contribute to the rise of global temperatures, and, subsequently catastrophic events. Therefore, a non-accidental relationship exists between human actions and climate change. Second, we know which human activities are responsible for driving carbon emissions, ways to cut them, and potential alternatives. So it is possible to establish a relationship between *tokens* (humans) and their effect on various aspects of the system through activities like air travel, meat consumption, energy consumption, and so on.

Thus, system consequentialist principles can be applied to examine human actions towards the environment because both epistemic conditions are fulfilled. In short, we possess

sufficient knowledge to make well-informed decisions to maximise systemic outcomes in this case.

Three tenets outline an argument for maximising systemic goodness (Hiller, 2016, 2013): the *Metaphysical Premise*, the *Axiological Premise*, and the *Better/Worse System Premise*.

The first and the *metaphysical premise* is teleological (Hiller, 2013). It is about an end and its relationship to a system: All human acts are means to an end[32], and the ends are such that they are a part of a system. So, in the case of climate change, all human behaviour occurs within a system. There is an relationship and interdependence between human ends and the environment. Environmental systems play an integral role in our sustenance. We need good-quality air to breathe, an optimal temperature to survive, food, and clean water to live. We also need fertile land for agriculture, trees for timber, and water bodies for irrigation, transport, and leisure. While our well-being depends on the well-being of the environment, it is a factual premise that the effects of human actions have thrown the climate systems off balance, as the evidence in Chapter 1 outlines. Thus, all human actions and the ends sought by those actions, such as food and energy choices, can be understood as parts of the ecological systems.

The second premise is called an *axiological premise* that refers to the non-instrumental goodness or badness of an action (Hiller, 2013). This premise prescribes that when an end is part of a system, the goodness of the system is not independent of the goodness of the end. It *can* be summarised as "For any x and y where x depends essentially

[32] Even the spontaneous human acts lead to certain ends, which have an impact on the systems (simply because they are happening within a system).

on y, if y is corrupted, so is x" (Hiller, 2014). To put it another way, if an end is corrupted, then the resulting system that constitutes the end is also compromised.

Consider another example of overfishing in a swamp to demonstrate this point. Fishing (the action) is a means to support the local livelihood (or y), and this activity takes place in a swamp ecosystem (x). To put it another way, the health of the swamp ecosystem (x) also depends on local livelihood (y). Overfishing, as a result of seeking y, can adversely affect the swamp ecosystem (x). Thus, while fishing is generally a means to a good human end (y), overfishing corrupts the well-being of a swamp ecosystem (x). In short, while fishing is a means to a morally valuable human end, its corruption (overfishing) negatively affects the ecological system in which it takes place.

In the context of climate change, this tenet adds an interesting proviso. Certain actions can be intrinsically good or bad. For instance, revisit the example of taking out a gas-guzzling SUV on a Sunday afternoon (1.5.3). These actions, even though non-instrumentally good by themselves (going out to spend quality time with one's family), are not independent of the systems they impact. In this case, taking the family out on a Sunday is a good end, but depending on how that end impacts the ecosystem, it can be morally impermissible from a system consequentialist view.

The third premise, or the *better/worse system premise*, suggests that there are better and worse ecosystemic states, depending on how humans act to achieve certain ends. In short, some ways in which humans act corrupt the ecological systems. So, in the swamp example, there can be two possible worlds. One in which humans seasonally move to other sources of livelihood, and the other in which they continue to overfish. The world in which humans control their fishing activities seasonally, allowing the fish population to thrive, is a better

world than the one where they do not. Therefore, there are better and worse human ends and actions, depending on how they contribute to a system.

This position has a significant advantage over other positions in environmental ethics because it clearly demonstrates why some intrinsically bad actions are prohibited[33]. For instance, positions by Singer (1975), Taylor (1981), or even Leopold (1949) are argued to allow a sort of 'environmental fascism' (Hiller, 2016). This means that a single moral principle, according to these positions, may absurdly allow the killing or mass displacement of humans to save other ecosystemic goods (Hiller, 2016). This problem is better dealt with by system consequentialism. A system consequentialist would note in this case that even if killing may uphold systemic well-being by reducing population, this, in itself, would not justify these actions because there are alternative ways that lead to a better system. For instance, long-term policy measures would uphold systemic goodness over sudden killing, displacement, or exclusion of groups.

Thus, Hiller's (2013, 2016) system consequentialist argument holds that when individual means to an end affect the larger human-nature or human-human systems, the moral goodness of an end depends on whether it enhances or degrades that system's goodness. So, all human actions, even those that are non-instrumentally beneficial to us, such as eating, commuting, and travelling, can be systemically good or bad depending on how they contribute to the ecological systems.

[33] Singer's (1975) sentientism and Taylor's (1981) strict egalitarianism may permit certain coercive actions, such as displacing a large number of people to save a sentient wildlife species from going extinct. If given an option between, say, displacing people or implementing family planning and education programs, these frameworks do not tell us which of the two would be better. Similarly, by focusing on ecological integrity, Leopold's (1949) version of ecocentrism may overlook other important consequentialist values, such as human autonomy. One may object that these positions would, much like system consequentialism, prefer the alternate option of family planning and education. Although this is true, I think they are charged with environmental fascism because they necessarily do not ask for an alternative, a better strategy that can uphold the systemic good. The principle that considers the possibility of a better system makes system consequentialism a more favourable position against these existing theories.

2.4.2. How System Consequentialism combines with Objective List

As already noted at the beginning of this subsection, we ought to subscribe to certain objective ends to respond effectively to climate change. We need two necessary lists of objective goods for maximising systemic outcomes. The first list has certain intrinsic, objective goods that are beneficial for ecosystems. We can call them ecosystemic goods. These are the kinds of consequentialist goods that ecocentrists desire: the flourishing of all beings on the planet and their relationships with one another that are fundamental for ecological balance. These ecosystemic goods can also be considered in the justification of other actions according to environmental system consequentialism.

The second list involves some objective goods that can maintain the balance of the ecosystem with other human-made systems. The second list is essential because it can ensure that all vital systems can co-exist with the ecosystems. As we examine the ethics of GN, Chapter 5 will also outline the second list. It will consider the ecosystemic benefits that are possible through GN, but it will also consider other values that are at stake in other systems critical to humans[34]. This will ensure that, in using GN ethically, we can maximise good ecosystemic outcomes while preserving the conditions for other systems (social, political, and economic) to flourish.

As it places value on consequences that one may not desire or express/recognise the desires of, but are still important for moral goodness, this position is a form of perfectionist consequentialist position. This is also a pluralist consequentialist position because it holds that there can be multiple things that can be used to measure the value of outcomes, or the 'best state of affairs'. This position does not prescribe a limited set of systemic values or goods, but instead, considers all the values that lead to good systemic outcomes as

[34] So, the second set of objective values can vary depending on the type of intervention we are considering to address climate change. As we will see, GN should uphold the value of prudence in the nudge (Chapter 5). When examining alternative environmental tools, such as carbon taxes, one might consider more specific concepts, like financial prudence or economic distributive justice.

consequentialist measures. In line with Avram Hiller (2011, 2013), who first formulated the theory of system consequentialism, these values/goods are important because they are necessary for achieving good consequences in a system.

This position requires a set of objective consequentialist values. However, it is different from objective list consequentialist theories because it explicitly focuses on the best state of affairs for a 'system'. A focus on a system is inspired by various ecocentric theories in environmental ethics, which claim the moral value of well-being within a system, seeking the flourishing of humans, non-humans, their interdependencies, and the relationships between them. Essentially, this position, *Environmental System Consequentialism,* adapts an ecocentric vision into a fully consequentialist theory.

2.4.3. How System Consequentialism combines with Ecocentrism

To maximise the good consequences associated with one's actions, we should claim a consequentialist position grounded in systems-based thinking. Such a position provides the theoretical basis for a consequentialist defence of GN[35]. A systems-based account of GN shows whose interests and well-being ought to be considered when acknowledging third-party harms imposed due to individual behaviour. This approach is based on a non-anthropocentric foundation for an environmental ethic; it aligns with ecocentric environmental philosophers, such as Arne Næss (2005), Holmes Rolston III (2011), and Aldo Leopold (1949).

In Chapter 1, problems with certain non-anthropocentric individualistic theories, namely, Singer's (1975) sentientism, Attfield's (2003, 2012) inegalitarian biocentrism, and Taylor's (1986) egalitarian biocentrism, were discussed (in 1.4). For both Taylor (1981) and Attfield (2003), having a life is the basis for moral worth. For Singer (1975), the ability to

[35] This provides a more comprehensive account of the ethics of GN than the positions that defend it as a basis for preventing third-party harm, such as Schubert's (2017b) and DesRoches et al. 's (2023) accounts. These accounts are discussed in Chapter 4 (A Literature Review on Ethics of Green Nudges).

feel pleasure or pain provides moral worth. These positions are individualistic. They focus on the inherent value of a being, failing to grant moral value to the biological interdependencies and relationships that underlie processes such as growth, survival, reproduction, and death.

An individualistic approach to address climate change is ill-suited because the consequences of our climate-related behaviour influence complex environmental systems. As outlined in Chapter 1, the effects of climate change are unequally distributed, manifesting through long-term disruptive changes and cascading effects that are difficult to predict. Individualistic non-anthropocentric positions focus on the moral value of discrete beings. They ignore the relationships, interconnectedness, and interactions between different parts of nature that are fundamental for the well-being of all beings. Thus, considering the best possible outcomes is not possible without considering the systems that are fundamental for environmental well-being.

Ecological sciences have shown how organisms exist in a symbiotic relationship, where the well-being of other living and non-living communities is vital for the survival of each (Leopold, 1949). Ecocentrists insist that when we ignore the conditions or the systems required for the survival of a being, including humans, we fail to extend our "ethical sequence" to the environment (Leopold, 1949).

We need a relational ethic–one that recognises how living beings and their relationships with non-living things in nature are necessary for the survival of the ecosystems. Such a line of environmental thought represents an ecocentric approach: it considers the moral worth of ecosystems as a whole, emphasises the intrinsic value of all beings, views them as interconnected and interdependent, and considers their ecological functions as

morally valuable. Three notable ecocentric theories are the Ecological Ethic (by Rolston III (1975)), the Land Ethic (by Leopold (1949)), and Deep Ecology[36] (by Næss (2005)).

An ecocentric position provides an adequate response to the question of whom do we value in assessing the best state of affairs. Such a position is scientifically astute[37]; we do not exist in isolation, and ecological functions are critical for the well-being of every living organism. Any beneficial consequences or state of affairs associated with an individual are not independent of the ecological systems that support their well-being. Such a position acknowledges the complexities in environmental systems. This means that any reference to maximising good consequences or the best state of affairs cannot adequately treat the effects of one's actions in isolation; actions have systemic consequences.

A system consequentialist position is an attempt to frame ecocentric views, whether Deep Ecology or the Land Ethic, within a fully consequentialist framework. It culminates from a general theory about the good consequences of a system, of which environmental ethics can be a subset. In what follows, the main tenets of ESC are outlined.

---

[36] Deep Ecology, as Næss (2005) frames it, is a mode of questioning that seeks to shift our fundamental understanding of ourselves, including matters of life and death. We are not separate from nature: acknowledging this and acting upon it is possible only through "self-realisation". In the process of self-realisation, Næss argues, we need to reject human domination over other life forms, increase our sensitivity towards the environment, promote eco-activism through art and culture, drive shifts in economic and policy reforms, and ultimately challenge the underlying value systems as we understand them. This is the strongest version of ecocentrism, and the rejection of 'separateness' between human and nature in deep ecology may challenge ESL on grounds that it prioritises/considers human goods over ecological goods.

[37] Rolston III (1975) argues that "Environmental sciences", which provide a better understanding of how humans function in a natural environment, their interdependence with the other parts of nature, is central to an ecocentric ethic. An ethic that deals with man's relationship to the land, according to Leopold (1949), must integrate individuals as community members with the land and the plants and animals inhabiting it. Callicott (1989) adapted Leopold's (1949) ethics into a consequentialist framework, arguing that moral consideration must be extended to ecosystems as a whole, rather than each member of the community. He appeals to the pattern of ecological "communitarianism," arguing that our moral concerns have expanded over time, from the self to family, to tribe, to communities, to countries, and so on. Such historical expansion of our moral concerns, he argues, can be extended to incorporate ecological systems.

2.4.4. A Newer Version: Environmental System Consequentialism:

In what follows, a newer version of system consequentialism will be outlined. Although it is a type of 'Rule-based Objective List System Consequentialism', in the context of addressing the climate crisis, it will be referred to as 'Environmental System Consequentialism (ESC)'.

Two modifications are introduced in Hiller's (2013, 2016) theory. First, actions justified by ESC should be according to the rules that maximise good ecosystemic outcomes most of the time. This is an indirect consequentialist version of system consequentialism. Second, actions are justified by ESC insofar as they uphold certain objective values that are important for long-term ecosystemic flourishing, as well as for maintaining balance with other vital systems. This makes ESC a perfectionist consequentialist position.

Hiller's (2013) position is a direct form of consequentialism. In line with ecocentrists, system consequentialism holds that systems are valuable in their own right. Even though actions occurring within systems have non-instrumental goodness or badness (through the *axiological premise*), their ethical permissibility depends on how they contribute to the system (Hiller, 2016).

An indirect consequentialist position requires us to act according to the rules that generally lead to the best possible outcomes. Addressing climate change and developing potential pathways to address human behaviour towards the ecosystem requires us to act, most of the time, in ways that maximise good outcomes for the environment. This is a stronger system consequentialist position because it emphasises the patterns of behaviour and is more compatible in the policy contexts, both of which are possible by following rules that maximise systemic well-being.

Individuals can have patterns of behaviour that are environmentally harmful. Acting according to rules can minimise ecologically harmful behaviour by guiding consistent and

long-term actions. Recall from the previous chapter (1.5.3) that a consistent pattern of imposing climate harms through one's actions makes the need for fulfilling the moral obligation stronger. Having rules about such actions works well because it promises widespread adoption of behaviour and improved pro-environmental behaviour patterns. This is more promising than assessing the systemic consequences of our actions each time.

For policy scenarios, rule system consequentialism works because it requires individuals to engage in pro-environmental behaviour most of the time, while also placing constraints on policymakers to act according to the rules that maximise good outcomes for the environment. In the case of GN, it places requirements on the nudgers to follow a 'good nudging practice'.

This means that while we should focus on achieving behavioural transitions, nudgers ought to implement nudges in a manner that can be generalised. For instance, this would involve designing nudges in a way that respects values such as autonomy and fairness (outlined in Chapters 4 and 6). Ethically, following rules about trust also has long-term implications for achieving good ecosystemic outcomes, a proposal explored in Chapter 6.

In practice, following the rules will involve nudgers acting according to a certain code of conduct while influencing behaviours. This will include sending feedback, engaging nudgees in the practice of designing nudges, and combining nudging with education campaigns. Moreover, another way to develop rules about GN is through Behavioural Self-Governance, a proposal developed in Chapter 7. All these ways of using nudges can be rules that maximise the long-term effectiveness of behaviour change, while upholding values necessary for other systems (social, economic, and democratic) to function well.

In short, adopting rules that lead to good systemic consequences, maximising the long-term benefits and flourishing of ecosystems, is desirable. This means that GN should

embody rules that, when adopted on a large scale, can lead to climate change mitigation. Thus, in this sense, indirect consequentialism offers a stronger ethical position for maximising good ecosystemic consequences. Moreover, such a framework enables GN to integrate within legal contexts and become part of institutional strategies to address climate change, preserving other connected systems as well.

ESC also requires actions to uphold, preserve, or at least not violate certain objective values. These conditions make the *axiological premise* by Hiller (2013, 2016) stronger as it restricts the compromise of some important human consequentialist values. So, instead of arguing that a system is better where humans are educated rather than mass-displaced, ESC can simply say that mass displacement is not an option because it violates the objective value of autonomy.

Having some objective values is necessary to uphold long-term ecosystemic well-being. This claim will be further explored in the next chapter, where it will be demonstrated how humans possess certain psychological traits that limit their understanding of the climate crisis and render their efforts to mitigate it short-sighted. While individual experiences of the system play a role in determining whether a system is flourishing or not, it is not the only basis for doing so. Since ecological systems comprise beings whose preferences are largely inaccessible or unknowable to us, preference satisfaction cannot be a sufficient basis for assessing systemic wellbeing. Therefore, there are epistemic limits to subjective measures of desire–they cannot capture what it means for a system (as a whole to flourish).

There is a set of values fundamental to well-being. It is better that we recognise these values, but even when we cannot (due to limitations of our preferences or experiences of pleasure/pain), they exist as the objective consequentialist goods that are necessary for the best state of affairs. These values may not be directly perceivable through experiences but are important for climate-related well-being. Thus, we should have an objective list of values that

are necessary for any system to flourish. These values may involve trust, fairness, prudence, principled autonomy, and ecosystemic health. These values are required for ecosystemic well-being. Moreover, having an idea about objective goodness also allows us to make better rules about a system.

Similarly, apart from ecosystems, there are other important systems that need to flourish because they are valuable in themselves. For instance, a social system or an economic system. These systems are valuable by themselves. Our actions occur within a complex, interlinked, and interdependent larger system, of which environmental systems are an integral part. Our actions must not undermine the goodness of other systems. Thus, starting with an objective set of values necessary for systemic flourishing can enable us to create rules and norms that preserve the long-term well-being of other systems, while seeking ecosystemic goods.

Thus, to summarise the newer version of this system consequentialist position: We must follow rules about a system that, when adopted, lead to systemic goodness, where systemic well-being also comprises some objective values. This means that the ecosystem has specific objective values that are essential for its long-term integrity, flourishing, and consistency with other systems. In other words, *Environmental System Consequentialism* prescribes that when given two possible actions, we must choose the one that follows a rule that prioritises the ecosystemic good, while upholding certain objective systemic values. Doing so is necessary to achieve good outcomes for the ecosystem, which is the ultimate consequentialist measure of value in this context.

2.5. Conclusion:

This chapter argues for a position in environmental ethics that provides a consequentialist basis for responding to the climate crisis. As this dissertation aims to

examine GN, which are essentially tools that address the climate crisis, this chapter has outlined the maxim according to which an action can be ethically justified in response to climate change. This means that we can now explore the consequentialist goods that ought to be aimed for by GN, as well as the objective values and rules that can make GN ethically justified. This endeavour is taken up in Chapter 5.

The next chapter shows how GN is an effective tool for responding to the climate crisis. This is important because, for GN to be justified from a consequentialist perspective, it ought to deliver the morally desirable outcomes. Two traits of human moral psychology that hinder pro-environmental outcomes, informed by Singer's (1972) discussion on limitations of human morality and the moral enhancement debate, are outlined. Then, it is shown how GN can help individuals act in response to climate change (and fulfil their moral obligations), irrespective of these two hurdles.

Chapter 3. Can Green Nudges help us fulfil our Moral Obligations?

As discussed in Chapter 1, behaviour changes are urgently required for addressing climate change–this is both a practical requirement[38] and a moral obligation. Traditional strategies like scientific warnings, moral suasion, and classroom-based learning have achieved success to a limited degree because human cognitive systems are not equipped to understand climate change as a slow-moving process (Kaaronen, 2018; Savulescu & Persson, 2015). For instance, when people eat meat, drive more, or frequently travel by air, it is difficult for them to grasp how their behaviour contributes to climate-related harms. Green Nudging (GN), at this point, is effective in helping us act according to our moral obligations by working with human cognitive systems. This is the main argument of this chapter.

Peter Singer's (1972) essay, *Famine, Affluence, and Morality*, considers a case similar to this one. Singer (1972) points out how people fail to act as their morality would prescribe when those who need our help are far away or unrelated to us. He also notes that we find it challenging to act with moral concerns towards a large number of people suffering in different parts of the world. These psychological traits have an evolutionary basis, as also illustrated in moral enhancement debates (Persson & Savulescu, 2015, 2008; Singer, 1981).

As the theory of ESC from the previous chapter outlines, we ought to act in accordance with the rules that uphold ecosystemic wellbeing (amongst other things). Yet these traits: *Epistemic Myopia* and a *Diffused Sense of Moral Obligation,* underlie our insufficient response to climate change. Recognising these as both a practical and an ethical challenge, this chapter considers how GN can align human psychology and behaviour in the face of climate change by working with these traits of human cognitive systems.

This chapter is structured as follows. The next section (Section 3.1) introduces nudging resulting from a shift in understanding of human psychology and behaviour.

[38] See: Nielsen et al., 2021b and Ewert, 2019.

Building on this, this dissertation introduces GN, mapping its functionalities, scope, and the parties involved, namely, the nudgee and the nudger (3.2). This step is critical as it provides a descriptive overview of GN and how it works, setting up the study for an ethical examination of GN.

Following this, taking up the main argument of this chapter, the role of human moral psychology is considered from the moral enhancement debate. These psychological dispositions underlie our moral failure to act in response to climate change, thereby preventing us from fulfilling our moral obligations (3.3). Then, it is shown how GN can address these traits and help us fulfil our obligations, presenting an account of how green nudges are effective (3.4).

3.1. What is Nudging?

The term 'nudge' concept was popularised by Thaler and Sunstein (2008) in their book, *Nudge: Improving Decisions About Health, Wealth, and Happiness.* They define a nudge as "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. To count as a mere nudge, the intervention must be easy and cheap to avoid." (Thaler & Sunstein, 2008, p.06). Policymakers implement nudges to encourage people to make better choices in areas such as public health[39], finance[40], and the environment[41].

[39] Default sign-up option for organ donation on driver's license forms can significantly increase the rate of sign-up for organ donation (Thaler & Sunstein, 2008). Similarly, communicating social norms about hygienic behaviour also shows promise (Thaler & Sunstein, 2008).

[40] Nudging through the default sign-up can significantly increase the sign-up rate for pension plans for the employees (Thaler & Sunstein, 2008; Sunstein & Thaler, 2003; Madrian & Shea, 2000).

[41] A study used social and cultural norms to nudge people to keep their surroundings clean. The "Don't mess with Texas" anti-littering campaign used targeted nudging to appeal to their collective identity. The campaign resulted in a 70 per cent reduction in littering (Nagatsu, 2015).

To understand what makes a 'nudge' distinct from traditional interventions[42] targeted at changing behaviour, it is better to start with how human behaviour was historically perceived in academic scholarship, particularly prior to the rise of behavioural economics in the late 20th century (Congiu & Moscati, 2022). For a long time, economists used single economic theories to inform both normative and descriptive market mechanisms (Thaler, 2016). For instance, it was assumed that market actors always act rationally in favour of maximising profits or utility. Real-world observations – showing that people act in much less rational ways across economic contexts – called into question the validity of such a perception of human behaviour.

By the 1970s, robust experimental evidence demonstrated that human behaviour in the real world departed from traditional theories (Congiu & Moscati, 2022). The field of behavioural economics emerged from this paradigm shift in understanding human behaviour, notably marked by the introduction of prospect theory (Kahneman & Tversky, 1979) and subsequently, the concept of nudges.

Prospect theory laid the background on which the idea of 'nudge' was coined and popularised. According to this theory, several factors, such as the timing and the organisation of choices communicated to someone, can influence their decisions (Kahneman & Tversky, 1979). For instance, loss aversion refers to the cognitive state in which the pain of losing a sum of money is felt more intensely than the pleasure of gaining an equivalent amount (Kahneman & Tversky, 1979). There also exists a tendency to resist any change, known as the status quo bias (Kahneman & Tversky, 1979).

[42] There can be a number of approaches to changing behaviour. For instance, by imposing a legal requirement to do something or prohibiting certain behaviours. Thus, traditional interventions in this context refer to interventions like taxation, laws, or mandates, all of which are common practices to achieve desirable behavioural outcomes.

Kahneman later wrote a popular science book, *Thinking Fast and Slow* (2011) building on his work on prospect theory, which posits that humans exhibit two types of thinking processes: System 1, which is fast, automatic, and unconscious, and System 2, which is slow, deliberate, and conscious (Marchiori et al., 2017; Thaler, 2016; Thaler & Sunstein, 2008). Humans often rely on the use of System 1 to make decisions. In other words, a lot of choices made by an individual on a day-to-day basis involve automatic and unconscious systems, rather than deliberative processes. For instance, an agent is more likely to choose a healthier food option if she notices it instinctively at eye level, even if her initial preference was to eat junk food (Thaler & Sunstein, 2008).

The cognitive mechanisms noted in prospect theory were mobilised by Thaler and Sunstein (2008) to propose that it is possible to leverage System 1 processes[43] of humans to "move people in directions that will make their lives better" (Thaler & Sunstein, 2008). From this insight, the idea of nudging emerged–intentionally changing the decision-making context of an individual to make their lives better.

Nudging is generally understood as tapping into non-conscious processes. However, behaviour change can be introduced by combining both System 1 and System 2 processes. Even though the initial ideas of nudging involved leveraging System 1 processes, experts combine it with other strategies to improve behavioural outcomes and thus tap into System 2 processes as well. This insight is reflected in advancing strategies in behavioural sciences, such as boosts[44] and nudge plus[45], to achieve long-term behavioural transitions. The common

[43] Another psychologically informed definition of nudging describes it as "an umbrella term for deliberate and predictable methods of changing people's behavior…to activate nonconscious thought processes involved in human decision-making"(Marchiori et al., 2017, p 03).

[44] Boosts aim to improve a person's capabilities and knowledge about a particular behaviour by providing them with specialised feedback, motivation, and educational messages (DellaValle & Sareen, 2020).

[45] Banerjee and John (2021) introduced the concept of 'Nudge Plus'. A Nudge Plus intervention adds an element of reflection through a feedback loop to help a nudgee perceive the implications of their behaviour effectively (Banerjee & John, 2021).

thread among all these behavioural strategies is that they operate on the foundational idea of nudging from a psychological point, changing the decision-making context to influence the cognitive systems of an individual.

This thesis uses the term 'nudging' to examine the ethics of changing behaviour through strategies that leverage both System 1 and System 2 processes. This is because considering the potential of only System 1 processes is inadequate in addressing climate change, which is an urgent behavioural and moral crisis. Oliver (2013) argues that, in pursuit of sustainable transitions, strategies that work with a nudgee (or the person being influenced via behavioural tools) without engaging them with other educational or moral deliberative strategies may be unwarranted. Such practices of nudging without education and engagement risk compromising the nudgee's ability to form preferences and make well-informed decisions by themselves in light of a collective crisis (Bovens, 2009).

Nonetheless, a behavioural expert may argue that calling such strategies 'nudging' is not theoretically and practically accurate. Different types of behavioural interventions leveraging System 1 and System 2 use specific terms, such as 'budge', 'boosts', and 'sludge' (DellaValle & Sareen, 2020; Nova & Lades, 2022). This practice, however, has not yet trickled into the discourse in the ethical literature in this context[46]. In other words, interdisciplinary engagements with the issue of behaviour change practices have held onto the term and its relationship with markets, governments, and citizenry (Gumbert, 2019; Hansen, 2016).

This thesis retains the term 'nudging' to engage with the ethical literature because excluding the term would disconnect this thesis from central ethical debates in behavioural

[46] See Malecka, 2023; Button, 2018. Małecka (2023) criticises the oversimplification of behavioural patterns and the misattribution of causality to cognitive mechanisms as a critique specific to nudging, rather than to the broader toolbox of behavioural science. Similarly, in conceptualising the role of the 'nudging state', Button (2018) argues that the behavioural turn in policy risks bypassing citizen capacities and democratic institutions.

ethics. Nonetheless, it is recognised that a better reference can be evoked through 'behavioural interventions' or a similar term.

The potential of using both System 1 and System 2 processes to improve pro-environmental outcomes through GN is considered in the scope of this work. GN come under the category of pro-social nudging, which seeks collective goals.

3.2. What is Green Nudging?

Based on who benefits or should benefit from a nudging strategy, a nudge intervention can be classified as pro-self and pro-social (Congiu & Moscati, 2022). While pro-self[47] nudges primarily benefit the nudgee, pro-social nudging[48] is primarily aimed at improving the welfare of society (Congiu & Moscati, 2022). GN is categorised as pro-social nudges[49] because it constitutes the class of nudges that aims to promote environmentally benign behaviour (Schubert, 2017b).

Considering that a nudgee engaging in environmentally harmful behaviour imposes harm on others, GN is defined as, "... behavioural interventions that aim to reduce or eliminate environmentally mediated harms to a third party" (DesRoches et al., 2023). Since the nudgee also benefits, albeit indirectly, from a green nudge, GN is also classified as all-regarding nudges (Siipi & Koi, 2022).

We are interested in the types of nudges that constitute a broad range of behavioural interventions that may not be specifically referred to as 'GN' in practice but ultimately aim at

[47] Examples of pro-self nudges involve promoting healthy food options, signing up for retirement saving plans, and warning labels on products like tobacco and alcohol.
[48] Pro-social nudges can be organ donation nudges, road-safety nudges, and hygiene-based nudges. While the nudgee may benefit from these to some extent (such as environmental benefits, hygiene-based, or road safety), the primary goal remains to improve the welfare of society.
[49] While green nudges in this instance impart benefit to the environment, their formulation as pro-social can be contested. For instance, green nudges encouraging reduced meat consumption may impact the cattle farmers. Thus, in considering the ethical permissibility of green nudges, we ought to consider their distributional effects (extent of fairness) on vulnerable groups. This theme is taken up in the ethical defence of green nudges from a system consequentialist view (Chapter 5).

reducing or preventing environmentally mediated harms. This ranges from energy nudges, which specifically target energy saving as well as saving money, to certain health nudges that promote vegan options as well as healthier food (Allcott, 2011; Allcott & Rogers, 2014; Nykänen et al., 2022; Reinholdsson et al., 2023).

Nudges aimed at improving energy consumption can be seen as a subclass of GN. OPower's "Home-energy reports" sharing personalised feedback through algorithmic processes is an example of GN. This is because saving energy has a pro-social element to it: reducing energy consumption means a healthier planet, and other parties (as compared to the nudgee) benefit most from this behaviour (Congiu & Moscati, 2022; Kasperbauer, 2017). It is also driven by contemporary algorithmic technologies to drive sustainable behaviour, so it can also be seen as a subclass of smart nudges or digital nudges (Lembcke et al., 2020; Ranchordás, 2019).

Similarly, nudging eco-friendly food options through the canteen menu has also caught the attention of researchers working in this area (Nykänen et al., 2022; Reinholdsson et al., 2023). A nudge can promote healthy eating as well as have a pro-social element to it, making it an example of GN. Various nudging strategies targeted at waste management, influencing shopping behaviour, and transport habits can be covered under the category of GN, although they may not be explicitly referred to as such in the studies (Andersen et al., 2018; Daugbjerg et al., 2014; Loan & Balanay, 2023; Riggs, 2019).

With human cognitive systems, GN functions in at least three ways. They tap into self-image, social-image, or use the inertial tendencies through default choices (Schubert, 2017b). These three functions are largely seen as the main cognitive mechanisms underlying pro-environmental behaviour driven by nudges, as well as other behavioural insights (DesRoches et al., 2023; Schubert, 2017b).

First, GN capitalise on the cognitive desire to maintain a positive self-image (Schubert, 2017b). Denmark has implemented a state-accredited procedure for ecolabelling the products sold in supermarkets (Daugbjerg et al., 2014). Ecolabels promote eco-friendly shopping behaviour by increasing the salience of product features, with an intention to make nudgees more aware of the environmental impact of their purchase (Daugbjerg et al., 2014; Schubert, 2017b). They tap into self-image by appealing to one's altruistic tendencies, and nudgees engaging in pro-environmental behaviour in this way derive cognitive pleasure (DesRoches et al., 2023).

Second, since the desire to possess a positive social image in a group is an evolutionary feature, using social norms and peer comparison to tap into the social image is another way to promote eco-friendly behaviour (Nagatsu, 2015; Schubert, 2017b; Tajfel & Turner, 1985). For instance, GN works with the social image through the OPower reports, which share household energy data in comparison with other neighbours (Allcott, 2011). Third, GN also function through defaults.[50] An effective example of this is signing up renewable energy suppliers as a default option when there is little or no difference in cost compared to conventional energy suppliers (Sunstein & Reisch, 2014).

To summarise the discussion on GN so far, GN aims to reduce environmental harms imposed by an individual's behaviour. By functioning in this way, GN have a pro-social element attached to it, rather than seeking only the nudgee's benefit. Moreover, from a functional perspective, GN leverages self-image, social-image, and default mechanisms. GN

[50] There are at least three reasons why nudging through defaults works: inertia (the status quo bias), loss aversion, and implicit recommendation (Schubert, 2017b, p.18). Inertia factors in through defaults because people perceive default options as an established reference point, so they avoid changing it. Similarly, loss aversion works through this effect, as people think of changing default choices as losing something that they already have. Implicit recommendation frames a choice as viable: people being nudged may feel that their service provider, employer, government, or policymaker endorses the default choice, thus leveraging the implicit recommendation heuristic (Schubert, 2017b).

is defined for the context of this research by adapting Saghai's[51] (2013) definition of nudges, which makes the role of nudger explicit, and is modified to propose a working definition:

> *A green nudges B when A purposefully triggers B's cognitive processes by using B's self-image, social image, or uses default options to ultimately reduce the environmentally mediated harms caused due to B's choices.*

Keeping a broad term 'cognitive processes' allows the scope for discussion about System 1 and System 2 processes, broadening the scope and applicability of this research. For theoretical precision, the term 'shallow cognitive processes' or 'System 1' can be used whenever suitable.

3.2.1. Who is a Nudgee?

A nudgee is the individual whose decisions or choices are influenced through GN. A nudgee could be an economic[52] agent, a political[53] agent, or a moral agent. As we are thinking through systems, there are different systems, like the ecological systems, social, economic, and political systems. These systems are interdependent and interconnected with one another. While the thesis primarily focuses on the nudgee's role as a moral agent who ought to minimise climate harms, it is important to note that their role as a political or an economic agent is not exclusive of their role as a moral agent. Chapter 7 explores how green

[51] The role of the choice architect qua nudger in the definition needs to be clarified to emphasise the role of nudger in a good nudging practice. Saghai's definition of nudges provides a template for this: "A nudges B when A makes it more likely that B will Φ, primarily by triggering B's shallow cognitive processes, while A's influence preserves B's choice-set and is substantially non-controlling (i.e., preserves B's freedom of choice)." (Saghai, 2013, p.491). The term 'shallow cognitive processes' refers to the use of System 1 processes.

[52] In the OPower example, the GN strategy may also work because of the economic incentives associated with saving energy.

[53] Such framing of nudgees finds use in policy contexts. For instance, while choosing between two competitive energy providers, the state often sets the renewable energy option as the default energy supplier (Sunstein & Reisch, 2014). In such cases, as costs are not a factor, a nudgee complies with a nudge because they consider their role as a civic agent whose decisions have a bearing on social good. So they function as political agent in this case.

nudging can be better integrated with self-governance through the participation of nudgees as deliberative political agents, as well as moral agents.

3.2.2. Who is a Nudger?

Any group, public or private, that nudges is a nudger. In the literature on GN, the government is usually referred to as the nudger, especially in the context of pro-social behaviour (Kuyer & Gordijn, 2023; Thaler & Sunstein, 2008). For instance, in the study conducted on the Danish eco-labelling system discussed above, the government acted as the nudger by standardising the eco-labelling system to facilitate informed decision-making through nudges. Moreover, Ewert (2019) also argues that governments need to pursue their role actively as nudgers to encourage behaviour transitions.

While governments are usually seen as the nudgers, private actors also play an important role in deploying GN, and they operate in a context where they are synchronised with each other (Lenzi, 2019). In such contexts, Lenzi (2019) argues that using GN ethically requires preventing the private entities from "out-nudging" the public nudges. Such nudges by private actors are consumptive nudges, which are driven by profit-seeking motives rather than environmentally favourable outcomes (Braganza & Kapeller, 2024). Even though the primary goal is to pursue public good through nudges, as is evident, there is also a competition between private actors to push their interests and gather traction through consumptive nudging in the same context (Braganza & Kapeller, 2024; Lenzi, 2019).

This follows a need to introduce some form of regulation to prevent the effects of consumptive nudging from compromising the effects of genuine GN that target behavioural change for pro-environmental outcomes. While the next three chapters (4 to 6) do not much engage with this caveat, the recommendations in Chapter 7 are aimed at minimising the involvement of 'consumptive' nudges.

Irrespective of who deploys the nudge, the term nudger in the thesis refers to the agency that uses GN to ultimately reduce the environmentally mediated harms, restricting the scope of discussion about any other kind of nudges.

3.3. Human Psychology, Nudging, and Moral Enhancement:

Recall from Chapter 1 that climate change is, in part, a behavioural problem shaped by psychological dispositions which fail to recognise it as an urgent threat. This is not to discount the political and economic factors that play a massive role. As we are looking at climate change as a systemic issue driven by human-centred systems, human behaviours contributing to these human-centred systems play a crucial role in driving this crisis.

Certain evolutionary traits underlie human cognitive systems that have restricted our ability to recognise our obligations in practice and improve our behavioural response to climate change (Persson & Savulescu, 2015, 2017a; Savulescu & Persson, 2012). These traits include prioritising near-term benefits over long-term benefits, caring more for people we know and those who are closer to us, being incapable of empathising with a large number of people, perceiving infliction of direct harm as morally worse than our failure to prevent harm and understanding our responsibilities in proportion to how we perceive imposition of harms (Persson & Savulescu, 2015).

In light of such threats and technological advancements, moral enhancement debates argue that human nature is a 'work-in-progress' and improving human cognitive systems to address these factors is morally desirable[54] (Persson & Savulescu, 2008, 2015; Bostrom, 2009). Nudging can be seen as working with similar psychological factors, and the arguments

[54] Transhumanism advocates this view, arguing that we need to use our scientific and technological advancements responsibly to profoundly improve humans, physically, mentally, and morally, to deal with these threats (Bostrom, 2009).

from these debates are relevant here. In this context, GN works with System 1 and System 2 processes, rather than only System 1 processes.

Enhancement interventions attempt to improve specific human capacities rather than changing persons as a whole (Juengst & Moseley, 2019). Activities, such as drinking coffee, taking benzodiazepines for anxiety, or lifting weights, are performed to transform humans into a better version of themselves. While these are temporary enhancements, biomedical advancements also improve certain capabilities permanently, such as vision through laser eye surgeries or mobility by prosthetics. Applied to moral psychology, enhancement refers to modifying traits like empathy, trust, and altruism through biomedical, pharmacological, optogenetic, socio-cultural, or neurological interventions (Savulescu & Persson, 2012; Juengst & Moseley, 2019). Another definition states that "Moral enhancements aim to morally improve a person, for example, by increasing the frequency with which an individual does the right thing or acts from the right motives." (Forsberg & Douglas, 2021). Evidence shows, for example, that propranolol reduces implicit social bias by improving empathy levels in treatment groups (Savulescu & Persson, 2012). Similarly, the perceived degree of trustworthiness is positively associated with oxytocin levels (Savulescu & Persson, 2012). These debates take a consequentialist position: if enhancement reduces existential risks, then it is morally desirable, and indeed a moral obligation (Savulescu & Persson, 2015).

It would be interesting to note that, from this view, if GN can be effective in enabling nudgees to overcome or act despite these psychological factors, such as shortermism, selective empathy and altruism, it would be a form of moral enhancement. Even though Savulescu and Persson (2012) present nudging as a proposal for human moral enhancement, they do not develop it.

If GN can increase the frequency of one acting according to their environmental obligation (outlined in Chapter 1) by addressing these psychological traits outlined above,

they would be a form of moral enhancement. Nonetheless, the changes in human psychology through GN, as it is, would be more temporary. If Kaaronen's (2018) proposal for a cognitive shift in how we respond to and understand climate change is realised through GN, it would be closer to the notion of moral enhancement.

Thus, GN has some interesting parallels with moral enhancement: nudging through informational messages and norms taps into the evolutionary psychological traits, such as short-term gratification, lack of empathy with the suffering of unknown and far-away individuals, and understanding of the indirect harms imposed by individual behaviour. As we will see in the next subsection, GN can address these traits through changing decision-making contexts and combining them with reflective elements, feedback, or rewards.

From an ESC position, moral enhancement would be justified to reduce existential risks to an extent that it does not violate other systems. GN can, thus, be seen as a practical, temporary, and less restrictive form of moral enhancement for climate change. And if cognitive shifts are long-lasting and this can reshape human moral systems, then the obligation to use GN would be more interesting for the moral enhancement debate.

The next section considers how GN is effective by helping us overcome certain evolutionary factors in our cognitive systems. These psychological factors frame our inadequate response to the issue of climate change on an individual scale and thus underlie our shortcomings in fulfilling our obligation.

3.4. How are Green Nudges Effective?

This subsection shows how GN helps us (or can potentially enable us over the long term) to overcome our evolutionary response towards climate change, which undermines our obligation to act. In exploring the ethics of GN, such a proposal can help us take a step further to show why GN is a morally effective avenue in light of human obligation to address climate change, cognitive systems, and human behaviour.

To develop this discussion, two psychological shortcomings of human psychology discussed in Peter Singer's (1972) essay are considered as paradigmatic cases: *Epistemic Myopia* and *Diffused Sense of Moral Obligation,* discussed in subsections (3.4.1) and (3.4.2), respectively. These psychological limitations are explained in light of the human moral enhancement debate. Then, it is shown how these limitations undermine our obligation to address climate change. Following this, it is shown how GN can address these limitations in practice.

3.4.1. Epistemic Myopia

Epistemic Myopia refers to the psychological state of lacking moral concerns towards those who we do not know or are far away from us. Moral concerns like altruism[55] and justice[56] are observed to be more elevated towards those with whom one finds some kind of association–due to the culture, race, or geography (Tajfel & Turner, 1985; Singer, 1972; Schubert, 2017b; Persson & Savulescu, 2017a). One of the reasons why this state dominates human psychology is that we possess strong moral feelings towards the suffering of others when we directly observe their suffering or find some kind of connection with them (Singer, 1972; Kollmus & Agyeman, 2002). It is apt to term it 'Epistemic Myopia' because much like the near-sightedness issue, we have a psychological limitation in possessing moral concerns for those whose suffering we do not observe or relate to.

The evidence for this psychological trait is backed by research in evolutionary sciences and psychology. Studies in psychology suggest that it is easier for us to help those we know and those near us at geographical and temporal distances (Schubert et al., 2024).

[55] Altruism is a universal element of morality–we try to do what makes others happy, content, or free (Savulescu & Persson, 2008, 2012, 2017a). Altruism is, "... a more sophisticated reaction which presupposes awareness that another party, who is like oneself in possessing a mind or consciousness, is experiencing pain or suffering that is expressed by overt behaviour." (Savulescu & Persson, 2008, p.170).

[56] Similarly, a sense of justice involves an appeal to the tit-for-tat concept or refers to the idea that we wish to treat others fairly (Savulescu & Persson, 2008; Persson & Savulescu, 2017a).

Human moral psychology is limited to 'group-based morality', which is a conditional response for survival (Persson & Savulescu, 2017a; Tajfel & Turner, 1985). Group-based morality suggests that humans evolved to care for smaller groups because it gave them a competitive advantage during the hunter-gatherer phase by maximising cooperation in small groups (Persson & Savulescu, 2017b). Similarly, the "exposure effect" suggests that we exhibit greater moral concerns for people we regularly interact with (Persson & Savulescu, 2017a, p.08). A sense of familiarity (kinship, geographical proximity, or relatability) plays a role in our moral concerns (including but not limited to the feelings of altruism, justice, empathy, and kindness) towards one another. Likewise, the suffering of a single person easily triggers our feelings of benevolence, sympathy, and justice compared to the suffering of a large group in which no individual stands out (Singer, 1980; Savulescu and Persson, 2017b).

This effect can be observed in relation to our environmental behaviour. We lack the motivation to help those worst hit by the effects of climate change due to a lack of familiarity with them or the observation of their suffering (Kollmuss & Agyeman, 2002). Thus, the result is that most of us have little motivation to address climate change despite being aware of the fact that changing our behaviours can help address the issue (Kollmuss & Agyeman, 2002).

While our psychological traits may limit the way we perceive our moral obligations, in reality, our moral obligation to address this crisis stands irrespective of whether we perceive it as an obligation or not. We think it is easier to help those who we know or who are near to us. But Singer writes, "It makes no moral difference whether the person I can help is a neighbour's child ten yards from me or a Bengali whose name I shall never know, ten thousand miles away" (Singer, 1972, pp. 231-2). Suffering caused by our environmental behaviour to someone far away or unknown is as bad as the suffering caused by us to someone near us. Causing harm is morally blameworthy irrespective of identity or distance.

Even if people in a position to significantly cut down emissions due to their behaviour do not witness climate-harm imposed due to their actions, they are still responsible for their actions as their actions form a link in the bigger system that causes suffering.

And it is in our capacity to address this kind of suffering. The descriptive fact that we do not see the utility of our actions by ourselves does not change the moral fact that there is a scope to minimise the climate-related harm through our actions. There is evidence that direct and indirect effects of human behaviour can have far-reaching consequences for the environment (Nielsen et al., 2021a). This means our susceptibility to Epistemic Myopia undermines our obligation to uphold ecosystemic good, because this psychological trait decreases our chances of feeling morally concerned enough to change our behaviour to address climate change. GN is one of the candidates to address this.

GN can drive eco-friendly actions by arranging the choices in a way that directs one's attention to their environmental impact, and thus leads to morally enhanced outcomes, such as, by increasing information salience and providing feedback. GN frames the impact of our actions in a tangible way, circumventing the effect of Epistemic Myopia in situ. Simply put, GN make it easier for us to fulfil our environmental obligations in practice, even when it is not possible for us to fully comprehend how our actions are contributing to climate change.

One practical reason why this is necessary is that environmentally relevant information is often organised in a way that comes across as complex to the nudgee, who is functioning with limited cognitive and temporal resources, as observed by Lehner et al. (2016). Being overloaded with information or choices leads one to rely on their heuristics. As a result, one acts in ways that undermine their obligations. To address this issue, GN can simplify complex information and make it easier for the nudgee to access environmentally-friendly options. Three examples of GN illustrate this possibility.

The first way is by increasing information salience and minimising cognitive dissonance through eco-labels. As noted previously, in Denmark, a standard eco-labelling system informs the nudgees whether products are eco-friendly, ethically-sourced, healthier, and better tasting (Daugbjerg et al., 2014). By aligning one's choices with one's environmental values, eco-labels help reduce cognitive dissonance and encourage behaviours that contribute to environmental goals.

The second way is visualising one's environmental impacts with the help of digital environments, which creates awareness about environmental goals and culminates in a better understanding of how personal actions contribute to climate change. Such initiatives may involve installing artificially intelligent smart meters or grids, which can provide insight into one's consumption habits (Sunstein and Reisch, 2014). Smart meters can examine nudgee's data and provide tailored recommendations to cut down energy usage. The third way is to use goal-setting nudges, which prompt individuals to commit towards eco-friendly behaviour in the future and track their progress according to their plans and commitment (Carlsson et al., 2021). An integration of digital environments to drive eco-friendly behaviour in daily aspects of one's life and goal setting within societal contexts can generate long-lasting habits and increase the awareness of how our actions can have a climate-impact.

These examples illustrate how GN can make the impacts of one's actions visible to them, ultimately making it easier for the nudgees to do the right action. They can also, over time, help us change our perceived understanding of our contribution to climate change, transcending the cognitive framing that underlies our failure to respond to this collective crisis effectively. This functionality of GN, in line with the definition of moral enhancement by Forsberg & Douglas (2021), attempts to increase the frequency with which nudgees fulfil their environmental obligations.

3.4.2. Diffused Sense of Moral Obligation

Another hurdle, *Diffused Sense of Moral Obligation,* refers to the psychological disposition of feeling that one's moral obligation lessens in the case of a collective action problem like climate change. There are two reasons why this psychological state is prevalent in humans.

First, we understand our obligation to address climate change only to the extent of our understanding of our contribution to the problem. This is to say, our conception of responsibility is causally informed, a tendency known as 'act-omission doctrine' (Persson & Savulescu, 2017a, p.10). When there is no visible causal link between the individual's behaviour and the environmental harm that occurs, it diminishes the guilt for not acting. Moreover, this effect is elevated when others around us evade their obligations (Singer, 1972).

Second, near-term bias–the cognitive tendency to prioritise short-term benefits over future consequences. This feeling is motivated by "a bias towards the near future", the bias to take care of immediate interests, and overlooking the benefit of acting in the remote future (Persson & Savulescu 2017b, p.231). Near-term bias in humans suggests that we are more concerned with immediate (or near-term) gains or losses over long-term consequences. Thus, when it comes to environmental actions, we carry the psychological trait of seeking short-term goals over long-term benefits.

Savulescu and Persson (2017b) observe that this bias is hard-wired into human psychology. While these psychological traits allowed the primates to tackle immediate threats in a brutal environment, they act as hurdles in addressing long-term issues like global warming and poverty. Together, these two factors lead to a psychological state of understanding our obligation to address climate change as diffused or weakened.

Although we feel that our moral obligation to act in the face of collective action problems is diffused or we feel less guilty, our obligation to address climate change still stands. In reference to individual behaviour in light of climate change, Garvey (2011) writes, "Given the enormity and complexity of the planet's climate system, it's hard to see how a single green choice, even a whole green life, could make the slightest difference to the suffering ahead." (p.161).

Our actions, however small, are scientifically linked to climate change. When we overlook the demand to act in eco-friendly ways, we become complicit in a lifestyle or a series of actions that cause harm or suffering to others. Thus, each contribution is valuable in its own sense, even if, as it appears to us, those actions do not seem to do much to address the problem. This is because the psychological state of feeling less guilty does not make us less guilty. As Chapter 1 outlined, we have a moral obligation to address it as our actions are causal links to climate change. So our obligation requires us to act to uphold ecosystemic goods, even though traits embedded in our cognitive systems prevent us from acting.

If GN can help us feel accountable for harm imposed due to our behaviour, point out the ways in which we can reduce our carbon emissions, and make us more aware of our actions, it can subdue the effects of the psychological state of *Diffused Sense of Moral Obligation.* In practice, GN can subdue the effects of this psychological trait by emphasising individual impact within the collective impact. To do so, GN can offer peer comparisons and personalise one's environmental impact to evoke moral concerns. Two strategies of GN are notable in this context. The first is "social proofing", which explicates one's environmental behaviour in comparison to others. The second one is "personalisation", which improves the perceived effectiveness of one's environmental actions.

GN that compares one's environmental impact with neighbouring households can significantly influence consumption patterns (Schultz et al., 2007). Described as "one of the

most notable non-price energy conservation programs", the nudging strategy, in this instance, presented data to consumers in a way that made them see their consumption with others, triggering peer competition (Allcott, 2011, p.1082; Schultz et al., 2007). This study uses social proofing–the psychological fact that people are influenced by the behaviour of others around them (or possess herd tendencies). By making individual behaviour visible in a collective context[57], the following experiment demonstrates how social proofing reduces a diffused sense of responsibility by emphasising individual responsibility to address climate change.

Similarly, on digital platforms, GN can help the nudgees see the impact of their actions by personalising it. For instance, a study conducted on Canadian participants used personalised recommendations in two avenues: driving and flying (Luo et al., 2023). Randomised controlled trials conducted on about a thousand participants showed that giving personalised recommendations and framing them in a comprehensive way for the nudgees, irrespective of the nudgee's background, gives promising results in reducing carbon emissions (with initial reports indicating 12 per cent reduced emissions within one year) (Luo et al., 2023). The study noted that influencing attentiveness about their choices and giving reminders about goals (memory boosting) has a positive potential for long-term change, as the nudgees can help them identify and understand the main causal drivers of carbon emissions in their lifestyle.

When effective, such personalised nudges can appeal to the self-image of the nudgee, driving them to strive for consistency between their perceived environmental obligations and actions. Similarly, on social media platforms, personalisation can drive the nudgees to see

[57] People have a diminished sense of environmental obligation because they see individuals around them evading their obligations. This makes it easier for one to not feel guilty for not acting in an eco-friendly manner. A nudging program like this allows one to see how their behaviour stands with the behaviour of others, helping them see how others around them are fulfilling their obligations.

how others around them are acting in climate-conscious ways, inducing a mechanism similar to social proofing.

Over time, this may have the potential to elevate one's environmental concerns and improve the capacity to make better moral judgements related to climate change over time. Thus, by specifying and communicating one's impact through social proofing and personalisation, GN can elevate the sense of environmental obligation and work like moral enhancement interventions.

There is some overlap between these two psychological traits. For instance, if individuals fail to see results from their environmentally conscious actions (*Epistemic Myopia*), they may believe their efforts are insignificant in addressing climate change (*Diffused Sense of Moral Obligation*). Similarly, when people empathise with those affected by climate change, they are less likely to feel that their environmental responsibility is diffused. Accordingly, certain GN strategies can play a role in addressing both issues. For instance, GN can help one learn about their actions through feedback and monitoring by drawing a relation between one's actions and collective impact (Allcott, 2011; Schultz et al., 2007). This overlap is not a problem. The goal is to show how GN can overcome the two psychological limitations, for which they show promise.

3.5. Conclusion:

This chapter outlined the descriptive framework of nudging and green nudging to supplement an examination of GN. It also made a case for GN as an effective strategy for mitigating climate change. This is because they promise a shift in behaviour, and (as we will see) in our attitudes, and our perceived understanding of our relationship with the environment. In other words, it was illustrated how GN can help us fulfil our obligations by working with evolutionary traits embedded in human cognitive systems.

Chapter 4. A Review of Ethical Debates on Green Nudging

This chapter focuses on the ethical issues and themes surrounding GN. Nudging is framed as a form of *libertarian paternalism* (Thaler & Sunstein, 2008). It is 'libertarian' in the sense that nudges respect the autonomy of the person being nudged. As people are free to opt out of choices promoted by nudges, nudging strategies retain freedom of choice and therefore respect autonomy. The claim for 'paternalism' holds that it is legitimate for policymakers to influence individual behaviour for improving personally and socially valuable outcomes. Thus, in being *libertarian paternalistic*, nudges influence the choices or behaviour of the person being nudged, while keeping the option to ignore the nudge open for the person being nudged.

This dissertation sets aside Thaler and Sunstein's political and philosophical framing of libertarian paternalism for two reasons. First, the use of the term has been theoretically contested in the literature on the ethics of nudges (Viale, 2019). Viale (2019) argues that this *libertarian* tag on nudges, even though it works theoretically, is less realistic. He writes, "The possibility of deliberative choice exists in abstract terms, but the choice architecture is biased…due to the prevalence of our System 1 over System 2" (Viale, 2019, p. 146). While it is true that people can decide to move away from default and fulfil their autonomy (which is the freedom to choose, from this view), in practice, they don't. This caveat, also pointed out by Małecka (2023), makes the applicability of the term 'libertarian paternalism' contentious.[58]

Second, conceptualising autonomy as freedom of choice is insufficient in the case of climate change. In dealing with the ecological crisis, individual freedom of choice should not

[58] Malecka (2023) has an interesting take on how System 1 and System 2 processes, defined by Kahneman (1972), are actually not separate mechanisms. They posit that such cognitive processes work in conjunction with one another.

be the main priority.[59] This point is expanded in Subsection (4.3.4) and then in (5.3.2), where it is argued that a different conceptualisation of autonomy, principled autonomy, is better suited in the context of climate change. Drawing on Onora O'Neill's (2002c, 2002d) work, principled autonomy requires individuals not only to have rights but also to fulfil universal obligations. From this view, when one's personal choices interfere with the well-being and rights of others (such as by contributing to climate harm), these choices are not fully autonomous. Therefore, in contexts where individual behaviour has collective consequences, we must prioritise moral obligation over freedom of choice.[60]

As the *libertarian paternalism* framing of nudging is both theoretically contested and ethically insufficient in the case of climate change, this dissertation focuses on the psychological functionality of GN, as has already been done in the previous chapter.

The purpose of this chapter is threefold. First, to review the existing ethical debates and questions surrounding GN. Second, to sketch the ethical concerns associated with GN against their potential to drive a meaningful change in our behavioural response towards climate change. This analysis helps to contextualise the main questions that this thesis seeks to address. Third, to consider whether certain ethical trade-offs can be considered insofar as GN helps us to fulfil individual obligations.

In the following section, a systematic review of the literature outlines the existing ethical concerns and themes associated with GN (Section 4.1). Following this, analysing the

[59]The reasoning for rejecting the view of autonomy as freedom of choice is threefold: First, climate-related harms imposed by individual behaviour is bad. Second, if autonomy is framed as "freedom of choice", then it allows people to keep making choices that contribute to climate harm. Then, it follows that autonomy as freedom to choose is not the moral positioning we ought to take in the case of climate change.

[60] Moreover, addressing climate change justifies some degree of paternalism. At least nudges aimed at pro-social behaviour (socially optimal behaviour outcomes) can be actively paternalistic and need not safeguard freedom to choose when one's behaviour imposes harm on others, a proposition that is already found in the literature on nudges, and is enunciated by Thaler and Sunstein themselves (DesRoches et al., 2023; Hilton et al., 2018; Oliver, 2013; Schubert, 2017b; Thaler & Sunstein, 2008). This point is also revisited in (4.3.1).

ethical permissibility of GN using a 'pro-tanto model' proposed by DesRoches et al. (2023) contextualises the main questions that provide the basis for a consequentialist defence of GN (Section 4.2). Lastly, it is considered whether certain trade-offs are justified considering the effectiveness of green nudges (Section 4.3).

4.1. A Review of the Literature on Ethics of Green Nudging:

The discussion around the ethics of GN is nascent, with only a small subset of the literature on the ethics of nudges exploring the topic. By illustrating the ethical concerns associated with GN, the following section places the challenges and concerns underlying GN in perspective. Through PRISMA-Methodology combined with Wohlin's backward snowballing method, this chapter identifies nine ethical themes that emerge in the literature on GN ethics (Kahrass (Knüppel) et al., 2021; Wohlin, 2014). Based on a thematic analysis of journal articles and book chapters obtained through the PRISMA search, nine ethical themes and issues related to GN are drawn from the literature. The systematic methodology of the literature search is available in Appendix A. Several ethical issues arise, many of which overlap with ethical issues with nudging more generally.

4.1.1. Third-Party Harm

The ethical themes and issues related to GN are seen in the light of their aim to mitigate harm to third parties. Climate change is a problem caused by externalities, which generates a rationale for economic interventions such as nudges (Kasperbauer, 2017). GN belongs to a separate category of nudges that use cognitive biases or moral concerns of citizens to influence actions that create negative externalities (Carlsson et al., 2021) or a market failure (Hilton et al., 2018; Schubert, 2017b). By encouraging individuals to engage in pro-environmental behaviour, "..they (GN) aim at encouraging people to voluntarily contribute to a public good, namely environmental protection" (Schubert, 2017b, p.331). GN, then, are targeted at reducing the third-party harm arising from climate change.

4.1.2. Paternalism[61]

As GN addresses the negative impact of one's actions on the environment, they also allow choice architects to decide what options benefit others, making 'paternalism' a significant concern. There is a spectrum of viewpoints regarding paternalism in GN. Paternalism is generally defined as "the interference of a state or an individual with another person, against their will, and defended or motivated by a claim that the person interfered with will be better off or protected from harm" (Dworkin, 2020).

On one end, GN is considered to be paternalistic due to their functionality (Hall, 2016; Hilton et al., 2018; Kasperbauer, 2017). The other side views them as non-paternalistic based on the goals (environmental protection) pursued by them (Santos Silva, 2022; Schubert, 2017b; Siipi & Koi, 2022).

Kasperbauer (2017) and Hilton et al. (2018) argue that, although GN is paternalistic like all other kinds of nudges, it is permissible because it addresses negative externalities and serves the collective welfare. GN aims to deter an individual from acting in a way that can potentially harm everyone dependent on the environment (Hilton et al., 2018; Ouvrard & Stenger, 2019). Moreover, some form of nudging is inevitable in energy regulation (Kasperbauer, 2017). This is to say that energy governance is an inherently paternalistic endeavour, and some degree of external regulation with nudges is required (Kasperbauer, 2017).

Hall agrees that GN is paternalistic (Hall, 2016) but suggests that this is in part the result of a narrow view of the self. She suggests that if one adopts a larger conception of self (viewing a community or a society as a self), the combined choices of a community harm the community itself.

[61] As discussed before, in the view of Thaler and Sunstein (2008), nudges are a form of soft paternalism. Choice architects do not block the available choice-set of the people being nudged. Instead, nudges attempt to make people better off by promoting choices that benefit them.

Siipi and Koi (2022) argue that GN is paternalistic if they promote an individual's welfare and non-paternalistic if they aim at social welfare[62] (Santos Silva, 2022; Schubert, 2017b). For instance, while energy-saving nudges benefit the person being nudged by saving them energy-related costs, they are still non-paternalistic because they primarily target environmental protection and energy saving. They argue that the justification of GN must be "categorically different from the libertarian paternalist justification of most nudges" (Siipi & Koi, 2022).

One major question is left open regarding paternalism in green nudges. It is about people who implement nudges or the choice architects. The obvious question is whether the choice architects have the necessary knowledge and integrity to nudge people (Hall, 2016; Lenzi, 2019). A potential response is that the presumption that policymakers are 'well-placed to make beneficent decisions' is common in public policy and is not specific to nudges (Hilton et al., 2018). Irrespectively, this presumption challenges us to consider the trustworthiness of the nudgers, such as their competencies and commitment to use GN ethically. This proposal is developed in Chapter 6.

4.1.3 Green Nudges in Markets and the Role of the State:

In analysing nudges as a sustainable market approach, Knoll (2019) notes five aspects of GN that are distinct from other sustainable market mechanisms, such as, Pay-for-success, cap-and-trade, and taxation[63]: (i) The loci of control can be a private or a public designer of a

[62] Siipi and Koi (2022) refer to 'climate nudges', the type of nudges that drive citizen behaviour towards climate-friendly alternatives. This is a terminological difference. Siipi and Koi (2022) talk of what I refer as green nudges.

[63] Pay-for-success refers to the state engaging with private actors, and allowing them to fund and manage a program aimed at achieving climate goals. The private actors are paid by the state only if the program is (assessed) successful. Cap-and-trade is another environmental policy tool that limits the extent to which a private entity can contribute to GHG emissions, and allows private parties to buy and sell (or trade) their emissions. Taxation or Pigouvian tax involves the state imposing penalties or fines on the polluters to discourage them from emitting and using the money for social costs (green initiatives and social programs). In contrast to all these, nudging, theoretically, does not involve any monetary aspect (Knoll, 2019).

nudge (which is referred to as the nudger in the dissertation); (ii) The facilitating agency of nudging interventions is a psychological and a paternalist state; (iii) There is no monetary control on the person being nudged (the nudgee); (iv) Nudgee is seen as an agent acting upon unconscious mechanisms and emotions; (v) Nudges are a result of post-liberal era, characterised by a shift away from fully market-based and regulatory governance toward behavioural governance (Thaler & Sunstein, 2008), powered by increasing digitisation (Knoll, 2019). These aspects of nudging have invited criticism on the role of the state in nudging and the perceived understanding of the cognitive heuristics and the capabilities of the nudgees.

The underlying idea behind nudges is that agents are emotionally driven individuals whose demands from the environment must be altered to fit with sustainable market mechanisms (Knoll, 2019). Nudging aims to transform individuals into better market participants through interventions by both states and corporations (Knoll, 2019). Like Knoll's view, Gumbert (2019a) points out that nudges for food consumption understand sustainability as a concept associated with the conduct of individual citizens. However, such an approach overlooks the public's susceptibility to corporate interests and behaviour regulation by market policies (Gumbert, 2019a).

Moreover, nudging might come across as off-putting to citizens, who are assumed to be making decisions on the basis of cognitive heuristics and covertness of nudges, leading to a decline in faith in the state (Lenzi, 2019). Lack of deliberation and persuasion on the part of the government and the use of GN might make the public feel alienated (Siipi & Koi, 2022; Kasperbauer, 2017).

The ethical implications of an intervention like an energy nudge for the state and the citizens can be assessed through three[64] criteria: information problem, multidimensional problem, and political economy problem (DellaValle & Sareen, 2020). Nudging can help address the information problem by providing individual and household feedback to the energy suppliers. Similarly, to address the multidimensional problem in the energy sector, nudging can help individuals overcome cognitive barriers by communicating norms and reminders to individuals or households about their energy use. Policymakers can overcome single-action biases and acquire the required information on energy behaviour using nudges to address the problem of political economy (DellaValle & Sareen, 2020).

4.1.4 Sustainability goals of Green Nudges

Several authors raise concerns about the sustainability goals associated with GN. Bornemann and Burger (2019) remark that the approaches facilitated by GN overlook the normative implications of sustainability. In their view, different conceptions of sustainability, namely, justice-based, resilience-based, and procedural concepts, are at play in the case of environmental regulation[65].

The relationship between nudging and sustainability is not a straightforward one because different ideas of sustainability challenge GN in various ways (Borneman & Burger, 2019). For instance, the procedural-concept might view nudging as technocratic and justice-based conception challenges issues of equity (DellaValle & Sareen, 2020; Hoberg & Strunz,

---

[64] Information problem refers to the idea that nudgees have limited knowledge about needs and problems associated with their energy behaviours. Multidimensional problem refers to the limited cognitive potential and motivation of individuals to use energy efficiently. The problem of political economy highlights the structural inequality between individual energy users and energy providers, where the nudgees have little to no ability to influence policy decisions (DellaValle and Sareen, 2020).

[65] Three conceptions of sustainability discussed by Bornemann and Burger (2019) are summarised:

a) Justice-Based Sustainability: Such a conception is based on ethical and moral investments that serve intra and intergenerational justice effectively.
b) Resilience-based conception of Sustainability: This group includes the conceptions of sustainability that involve maintenance and well-being of local and global ecosystems.
c) Procedural-conceptions: Unlike the former two conceptions of sustainability, procedural conception understands ‘Sustainability’ as a political concept that continuously evolves in societal discourse.

2018). Similarly, the resilience-based conception raises concerns about whether individual rationality can lead to collective rationality.

DellaValle and Sareen (2020) look at nudging as an intervention to address the problem of Energy Poverty[66]. If one looks at GN as a means to achieve justice-based goals (such as merit goods[67]) in the future, then it would be ethical to override the short-term preferences of individuals to satisfy their long-term preferences (Hoberg & Strunz, 2018). Further, Dobson (2014) contends that the ideas of 'sustainability citizenship' cannot get the required attention in the public contexts amidst the short-term policy-making strategies of the states. In his view, nudging *eschews* the normative debate and hides the goal of creating a "sustainable citizen" from view, and thus escapes accountability (Dobson, 2014, p.03). While governments are responsive bodies in debates surrounding sustainability, the discussions fall short of outlining their roles as responsible entities[68] while only focusing on improving the role of citizens as consumers (Dobson, 2014).

One way to address this issue is by bring nudgees at the forefront of designing, implementing, and regulating GN. This is possible through a bottom-up approach for assessing and receiving feedback on GN, an approach developed in Chapter 7.

---

[66] Energy Poverty is a condition where individuals are unable to secure adequate home energy services, a state that violates the principles of Energy Justice (DellaValle & Sareen, 2020).

[67] A justice-based idea of sustainability would consider intra and intergenerational justice as a merit good. Hoberg and Strunz (2018) define merit goods as the goods for which the social planner chooses to prescribe choices that are different from individual preferences. From such a view, one could take GN as a means for achieving environmental goals such as climate protection. A person cutting down her consumption due to GN would save fossil fuel for future generations.

[68] Dobson (2014) further argues for an Aristotelian sphere of politics where the public gains understanding by debating and enacting rather than avoiding deliberation and engagement. In the words of Dobson (2014), 'sustainability citizenship' requires that a citizen: (A) Believes sustainability to be a common good that is not achievable by merely seeking self-interests. It should be driven by public engagement, public participation, and fair distribution of environmental resources. (B) Driven by other-regarding interests along with self-interests. (C) Believes ethical and moral knowledge is necessary along with the pursuit of techno-scientific knowledge. (D) Believes that a sustainability right obliges citizens to carry a responsibility to behave sustainably. A way to introduce this concept into governance is through self-governance, taken up in Chapter 7.

4.1.5 Autonomy[69]

Individual autonomy is a common theme regarding the ethics of nudges. Unsurprisingly, autonomy emerged as a prominent topic in the literature on GN. There are three views regarding autonomy in GN. First, GN affect autonomy, but is justified considering the climate crisis. Second, a compromise on autonomy is unjustified as it leads to a fragmented preferential structure. Third, whether GN undermines autonomy depends on the conceptualisation of autonomy and the types of nudging techniques employed.

The first view suggests that a compromise on autonomy is inevitable and necessary to address climate change. According to Ouvrard and Stenger (2019), the argument that nudges do not respect the freedom of individuals 'holds no water' in the case of climate nudges. They contend that if an agent becomes incapable of living where he wishes to due to lack of environmental reasons, he will lose his freedom anyway (Ouvrard & Stenger, 2019). Moreover, individuals are already acting within a framework established by the government for energy consumption. Therefore, autonomy is always compromised while using energy (Kasperbauer, 2017).

Tobias Gumbert (2019b) holds a contrasting view, in which people choosing in the presence of a nudge do not make an autonomous decision on their own; instead, they conform to 'implicit external expectations' (Gumbert, 2019b, p.114). Furthermore, Gumbert (2019b) worries that non-autonomous decisions made in the presence of nudges can lead to incoherent preference structures in individuals, leading to long-term implications on prudence (Gumbert, 2019b).

The third perspective on autonomy in nudges raises two points. One can infer whether nudges compromise autonomy depending on the conceptualisation of autonomy adopted and

[69] The conception of autonomy as freedom of choice is by far the most adopted idea of autonomy in nudges. In the view of Hoven et al. (2020), autonomy as freedom of choice is compatible with Mill's idea (*On Liberty*) of individual liberty and harm principle. According to it, the liberty of individuals should not be constrained unless it is important to restrict their actions to protect others.

the type of nudge strategy used. While some conceptualisations view autonomy as an absolute concept (Hausman and Welch 2010; Bovens 2009), authors like Buss (2012) contend that as long as nudges enable someone to 'act per their character or conditions of minimal human flourishing, ' people act autonomously. According to Schubert (2017b), this view is promising. In short, GN preserves autonomy when people lack clear environmental preferences, helping them act according to their true preferences (Schubert, 2017b; Siipi & Koi, 2022). On this view, however, GN challenges the autonomy of climate change deniers (Siipi & Koi, 2022).

Furthermore, different kinds of nudging techniques challenge autonomy on various levels. Santos Silva (2022) compares nudging with self-nudging interventions or boosts, where the latter requires the active involvement of the recipient of the nudge. As they require cooperation and are explicit, self-nudges and boosts do not challenge the autonomy of people (Santos Silva, 2022).

Similarly, default GN challenges autonomy because they impose an implicitly endorsed choice rather than fostering the competence of the nudgee to make eco-friendly decisions by themselves in future (Schubert, 2017b). The concerns regarding autonomy in nudges overlap with the issues regarding transparency and manipulation. We will discuss them in the next section.

4.1.6. Manipulation due to lack of Transparency

There are also concerns over covert forms of GN (Santos Silva, 2022; Ouvrard & Stenger, 2019; Gumbert, 2019b). Nudges encourage an intentional change in the 'autonomous decision-making process', often in non-transparent ways, resulting in manipulation (Gumbert, 2019b; Santos Silva, 2022). There are at least two reasons why manipulation due to lack of transparency is concerning in nudging: status quo bias and sharing incorrect information.

Nudges that use status quo bias are manipulative because nudgee are unaware while being nudged (Ouvrard & Stenger, 2019). GN, such as the diffusion of information (carbon footprint of a product), does not manipulate agents as much as subliminal advertising manipulates them (nudges that use stimuli or olfactory senses). Explicit nudges do not compromise an individual's autonomy on the same level as other subliminal nudges (Ouvrard & Stenger, 2019).

Nudges that share false information are arguably manipulative. Hilton et al. (2018) give an example of a study (Schultz et al., 2008) in which incorrect information was communicated to the public through social norm nudges to influence their behaviours. If implemented, sharing false information through nudges could threaten to mislead the public and create a feeling of distrust against the authorities in charge of nudging (Hilton et al., 2018). While taxes and mandates are transparent, interventions like GN can be used cunningly in non-transparent ways, leading to deception (Siipi & Koi, 2022). Furthermore, non-transparent GN can lead citizens to believe that the state views them as agents prone to manipulation and persuasion (Siipi & Koi, 2022). Gumbert (2019b) recommends that choice architects must ensure that individuals willingly conform to nudges.

4.1.7. Privacy Issues

Use of GN in digital environments gives rise to privacy concerns. For instance, Sunstein & Reisch (2014) suggest that the kind of nudges that uses default options to nudge individuals should be personalised. Personalising GN that influences pro-environmental behaviour requires a high volume of statistical data on behavioural patterns. The need for data to personalise nudges generates an insatiable thirst for information on people's behaviour; this is why digitisation comes across as a natural ally of nudging (Knoll, 2019, p.250). In the current scenario, digital platforms are crucial in implementing nudges and other behavioural interventions.

Hilton et al. (2018) argue that privacy concerns associated with informational GN (that work by making options more salient or available than others) are less pressing. In contrast, GN that uses the personal data of individuals (goal setting, personalised defaults, monitoring consumption) raise concerns regarding privacy as they risk exposing personal information to third parties without consent (Hilton et al., 2018).

4.1.8. Fairness

There are also fairness concerns regarding GN. Nudging policies run the risk of allowing a minority to free-ride on the efforts of the majority (Lehner et al., 2016; Ouvrard & Spaeter, 2015). Like all policy interventions, different citizens respond to nudges depending on their social, cultural, economic, and personal backgrounds. "Political sensitivity, environmental vulnerability, and educational characteristics play an important role in the efficiency of policies based on nudges." (Ouvrard & Spaeter, 2015, p.54). For instance, using nudges such as green defaults in a trivial way can impose unfair distribution costs on the population (Sunstein and Reisch, 2014).

Schubert (2017b) argues that we can judge the extent of fairness of GN through its (1) redistributive impacts and (2) the risk posed by them in diverting attention from 'socio-institutional problems" that are at the heart of many societal problems, such as climate change. The problem of redistributive impacts concerns the heterogeneous response from the citizens (Ouvrard & Spaeter, 2015; Schubert, 2017b).

One of the most common examples of the heterogeneous effects of GN is moral licensing. Moral licensing involves people deviating from pro-environmental behaviour in certain areas because they are already exhibiting eco-friendly behaviour in some other aspect of their lives (Hilton et al. 2018). For instance, a person buying biodegradable detergent may no longer feel obligated to cut down her electricity consumption because she feels morally fulfilled by making pro-environmental choices in the supermarket. It is also possible that

some sub-populations can perceive nudges (feedback, social norm nudges) as a form of moral licensing and, as a result, they might end up showing contrasting behaviour (DellaValle & Sareen, 2020; Hilton et al., 2018).

4.1.9. Using Nudges for long-term behaviour change

Over the long term, GN can hamper the decision-making abilities of people being nudged (Hall, 2016). Contrary to this view, some authors see a radical potential in nudging to transform people into eco-citizens. They suggest that nudging should be used to develop moral motivation in citizens by changing the way people look at environmental issues. Given the urgency of the climate crisis, there are good reasons to take nudging seriously to address climate change since using behavioural influences is inevitable in environmental policy-making (Hukkinen, 2016). Moreover, nudging can free up one's mind to achieve more significant goals of societal transformations and collective action (Borneman & Burger, 2019).

An interesting perspective on the "philosophy of nudging" emerges in Kaaronen's (2018) work. Nudging is a potential tool to address the human-nature dichotomy–the perspective that humans are separate entities from the natural world. A 'processual approach' to the environment can help us understand slow-moving processes like climate change in a different light (Kaaronen, 2018, p.11) and help citizens transform their relationship with the environment. An example of this is receiving real-time impact on energy behaviour, allowing one to observe the "interconnectedness" of one's actions with the ecological consequences (Kaaronen, 2018).

Kaaronen interprets nudging as a "political means of designing cognitively attractive and empirically sound mental models that can alter people's behaviour towards socio-ecologically sustainable transitions" (Hukkinen, 2012, 02, as cited in Kaaronen, 2018). They suggest that nudging on a deeper cognitive level can help people recognise their

responsibility towards the environment, perceive the effects of behaviour changes in real-time, and see themselves as a part of the system that needs to be changed to mitigate the climate crisis.

Having reviewed the nine ethical themes identified in the literature, it is worth addressing a more structural critique of GN that contextualises these themes within a broader debate about individual versus systemic approaches to climate change. Chater and Lowenstein (2023) raise concerns about GN as category of interventions that focus on i-frames, working with individual behaviour, cognitive processes, and habits. Their argument is that GN targeting individuals rather than the systems within which individuals operate is insufficient and perhaps counterproductive. GN diverts focus from the s-frames, which incorporate systems of rules, norms, and institutional reforms (also see: Schubert, 2017b). They worry that i-frame interventions serve the interests of corporations seeking to resist systemic reforms by framing collective action issues as matter of individual responsibility. From this view, GN may be counterproductive.

This argument raises a more serious concern: GN targeting i-frames focus on narrow and concentrated outcomes. They also pose a hindrance to effective climate transition by affecting possibilities for systemic changes.

There is, however, a caveat to this situation. Chater and Lowenstein (2023) present economic and regulatory interventions as primary examples of s-frame interventions. Yet their definition of s-frames includes systems of norms, rules and institutions. Even though they acknowledge the intertwining of s-frame and i-frame interventions, they do not develop the idea. The norms that policymakers incorporate within top-down policies are driven by i-frame factors such as self-image and social signalling. Along with the norms, traditional regulatory interventions like carbon taxation are also driven by i-frame factors like social-comparison (Gravert & Shreedhar, 2022). This indicates that the boundary between the i-

frame and the s-frames is not just unclear, but that s-frame changes are also dependent on i-frame changes.

GN justified by ESC substantially address the concerns raised by Chater and Lowenstein (2023). ESC guided GN is justified through a normative framework that focuses on improving outcomes, but also the systems within which these outcomes are embedded. In assessing GN by this position, we are focusing on not only driving behavioural changes, but also virtuous circles and cognitive alignment (See 5.1.). Instead of assuming norms (which are s-frame changes) forming from top-down approaches, GN assessed through ESC works with the processes of norm formation– by cultivating improved habits, awareness, behavioural shifts in individuals. Such i-frame driven interventions that cultivate improved understanding of nudgees as eco-conscious citizens, as well as a better understanding of how one's actions contribute to climate change, underlie desired s-frame changes[70].

Chater and Lowenstein (2023) also worry that the current pattern of applying GN (which focuses on i-frames) risks crowding out s-frame considerations in policy contexts. However, the crowd-out issue is applicable to any policy measure. This is an implementation issue rather than an issue inherent with driving individual behaviour changes. GN justified through ESC, which are implemented through bottom-up governance mechanisms (See 7.3. and 7.4.) can support s-frame changes rather than undermining them.

4.2. How does Green Nudging raise Ethical Issues?

While GN aims to improve behaviours in response to climate change, which is a morally desirable end, it also imposes some unaccounted negative effects on the nudgee, as discussed above. However, like any intervention capable of bringing a positive change,

[70] Moreover, the i-frame and s-frame distinction is not as clear as Chater and Lowenstein (2023) suggest. Individual behaviour is embedded in systems and systems emerge from complex cumulative individual behaviour. The leverage for addressing climate change through s-frame changes lies at the interface of s-frame and i-frames, a position where GN can be advantageous.

potential negative consequences are unavoidable. Thus, it becomes necessary to weigh the benefits of GN against the harm imposed by it on the nudgee (DesRoches et al., 2023).

DesRoches et al. (2023)[71] examine the benefits of effective GN against its adverse effects on the nudgee through a 'pro-tanto' model to provide a "high-level normative guidance" to policymakers[72]. They write, "...a GN is ethically permissible when the intervention is efficacious, cost-effective, and has advantages (i.e., reducing the environmental harm to others) that are not outweighed by countervailing costs or harms for nudgees." (DesRoches et al., 2023, p. 1). They argue that in designing nudges, the nudgers must consider the welfare effects of a nudge through the model given below.

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

Figure 1: The Pro tanto Model by DesRoches et al. (2023).

This model maps the level of harm (*Ordinary* and *Catastrophic*) against the net welfare effect on the nudged person (or the nudgee) and the third party. The horizontal slots in the diagram show that an efficacious nudge can prevent ordinary or catastrophic harm, and impact a nudgee positively, negatively, or have no net welfare effect on the nudgee (shown by signs '+', '-', '0'). If a nudge impacts a nudgee negatively, their ethical permissibility must be examined (represented as red and a '-' sign).

[71] As this journal article was published after the systematic literature review of GN in April 2022 (See Appendix A), the discussions from this study are taken up in a separate subsection.
[72] In explaining the use of the term pro tanto in this context, the harm caused by individual actions provides a nudger with a reason to intervene and improve behavioural outcomes, unless there are other reasons that outweigh the benefits of preventing that harm.

Whether or not GN is permissible depends on whether it prevents a non-negligible or a moderate damage (ordinary harm), or severe, large-scale, and irreversible damage (catastrophic harm). So in cases where GN is preventing ordinary harm, its justification may require respecting ethical values like fairness and autonomy. Simply put, based on this model, the ethical permissibility of GN hinges on the seriousness of the harm it targets and the countervailing costs to the nudgee.

There are two reasons why anticipating ordinary or catastrophic harms due to the nudgee's actions is not unfounded. First, there is a possibility that our actions have cascading effects. There are certain thresholds to climate change, which if crossed, can lead to dramatic effects, causing extreme vulnerabilities, and a series of climate change-induced disasters driven through interdependencies between different ecosystems (Bowen et al., 2022; Lawrence et al., 2020). To avoid the risk of crossing such threshold points, individual actions play a valuable role (Broome, 2019).

Second, our climate impact does not occur in isolation and cannot be characterised only by individual carbon footprints. By engaging in certain behaviours, such as purchasing products that have adverse environmental impacts during production, taking flights more frequently, using personal transportation, and heating or cooling our homes excessively, we initiate a series of actions that inspire others and contribute to the global infrastructure that imposes climate stressors. This observation is supported by a study conducted by psychologists in Denmark, who note that individuals from a higher socioeconomic class not only function as consumers but also serve as role models and investors (Nielsen et al., 2021). Thus, by limiting such behaviours on an individual scale, GN plays a role in preventing ordinary or catastrophic harm[73] due to the nudgee's behaviour.

[73] Moreover, anticipation of such an extent of harm that can be ordinary or catastrophic is rational. By Precautionary Principle, if there is evidence that an activity may lead to harm, precautionary actions are

In short, for GN to be permissible, it needs to be (a) efficacious and cost-effective, and (b) the benefits of the GN (reducing third-party harm) are not outweighed by countervailing reasons (significant harms for nudgees) (DesRoches et al., 2023). For (a), ineffective nudges are considered to be unethical as they burden the resources (Schubert, 2017b; DesRoches et al., 2023). Ineffective nudges involve the nudges that lead to no positive effects for third parties. Similarly, although the authors do not specify what 'cost-effective' entails, it can be interpreted as the economic and social costs of implementing, designing, and regulating GN. This model provides a basis for exploring the ethical permissibility of GN by probing (b) through four questions:

I. What sort of harm is the nudgee susceptible to due to GN?
II. How does GN prevent harm?
III. How to weigh environmental benefits against the countervailing harms on the nudgee?
IV. How do we minimise the overall harm to the nudgee?

The review of ethical themes and issues associated with GN in Section (4.1) responds to (I) by outlining the countervailing ethical burdens on the nudgee, such as, loss of autonomy, paternalism, fairness, and so on. To address (II), an account of the ethics of GN, then, needs to illustrate whether and how GN is effective. The previous chapter responds to (II) by showing how GN works with the cognitive systems of the nudgees. A worry is that i-frame

necessary. This principle is often employed to guide decisions in risky and uncertain situations in law and policy (Christiansen, 2019). Andreas Christiansen (2019) lays out an argument defending Precautionary Principle (PP), arguing PP is as rational as Expected Utility Theory (EUT) when considering the possibility of high stakes environments, like a public health crisis, rise of Artificial General Intelligence, or Climate Change. The principle states that if we know a serious risk and a recommended precaution, then a demand for precautionary action is triggered (Christiansen, 2019). In the case of climate change, we possess sufficient evidence that individual behaviour has an environmental impact, and a well-weighed possibility that extensive carbon emissions can trigger catastrophic harm. Thus, using GN to prevent such harms due to individual behaviour can be reasonably justified through this principle.

interventions are perhaps insufficient to deal with the collective action problems because they yield “small or null results” (Chater and Lowenstein, 2023, p.06). However, a potential response is that the poor performance of GN is a reflection of design and implementation flaws, rather than an underlying limitation of behavioural interventions as a category. When justified through an ESC-based framework, GN can be integrated as s-frames interventions, to deal with the climate crisis.

Moreover, a study by DesRoches et al. (2026) also provides support for the ethical permissibility of GN via a decision-making framework rather than a consequentialist direction. The authors use a Lexical Precautionary Principle (LPP) to view GN in a much more serious light: as tools to address climate change as a catastrophic event. This positioning of GN allows even libertarian-leaning positions to acknowledge the importance of GN and other behaviourally-mediated interventions (DesRoches et al., 2026). There is convergence between this approach and ESC. While DesRoches et al. (2026) argue for justification of GN through the LPP approach, ESC arrives at a compatible judgement through a normative framework focusing on systemic values. The fact that different theoretical approaches support ethical use of GN supports the case for pursuing behavioural changes more actively in response to climate change.

Considering the positive consequences associated with GN, a comparison requires us to consider whether a trade-off on some ethical concerns associated with GN is possible in light of the urgency to address climate change (III). So in the next subsection, some ethical trade-offs are considered. Finally, to weigh the benefits against the harms due to GN through a systemic view, the next chapter (Chapter 5) applies the ESC developed in Chapter 2 to outline the ethical permissibility of GN.

Following this, Chapter 6 considers trust as a condition to minimise countervailing burdens on the nudgee (to address IV). Moreover, institutional trust in nudges can also lead to

a better state/citizen relationship between the nudgee and the nudger, leading to a better system of GN. In making this argument, Chapter 6 appeals to the ubiquity of trust using Annette Baier's (1986) scholarship, and then proposes an account of institutional trust between the nudgee and the nudger using Onora O'Neill's (2002f, 2018) theory on trust and trustworthiness.

4.3. Can we consider some ethical trade-offs if Green Nudging is promising?

This subsection explores whether the severity of the climate crisis suggests that we should overlook some of the ethical issues of GN discussed previously, if it means they are effective in minimising climate-related harms imposed due to one's behaviour.

4.3.1. Paternalism and Green Nudging

It is difficult to argue that GN is not paternalistic (at least to some degree), but this does not mean they cannot be ethically justified. Climate change is a problem, and the broadest justification of paternalism is simply consequentialist; paternalism is justifiable to prevent more harm (Dworkin, 2020). Some views discussed in (4.1.2) indicate that a trade-off regarding paternalism can be considered if it means climate harms can be minimised (Hilton et al., 2019; Siipi & Koi 2019; Thaler & Sunstein, 2008).

When examining the climate-harming tendencies of an individual, various paternalistic interventions, such as taxes, bans, or mandates, are actively implemented. This precedent suggests that the use of paternalistic interventions to address climate-harming tendencies is common in cases where one's behaviour has negative implications for others. Thus, given this context, paternalism is not a strong enough reason to refrain from using GN to improve environmental outcomes. Even if GN is paternalistic, it can be permissible when effective and shows promise for addressing the environmental crisis.

4.3.2. Fairness and Green Nudges

Nudges are targeted at heuristics and psychological processes; different types of GN strategies would be required to nudge a heterogeneous group of people with diverse social, cultural, and economic backgrounds. In the absence of this tailoring, different levels of awareness, knowledge, and background result in varying responses to nudges. This results in redistributive effects, that is, unfair distribution of benefits and harms administered by GN (Schubert, 2017b). With the technological means to tailor GN based on one's background, the issue of fairness of burdens and benefits of nudging is more closely tied to the way GN is implemented, rather than something inherently unfair about GN.

If nudging is implemented in a way that imposes little to no redistributive burdens on the nudgees, GN would be fair. Institutional trust can provide a way to address this, taken up in Chapter 6.

Another way in which nudges can be unfair is by distracting from structural problems that underlie climate change (Schubert, 2017b). If nudging can lead to the mental shift that helps us become more aware of our role in climate-related harms, we can identify the institutional changes necessary to mitigate climate change and advocate for them in democratic settings (see Chapter 5). By appealing to certain moral traits in us, such as altruism, a sense of justice, and empathy towards the suffering caused by climate change, GN can help us identify institutional barriers at the heart of climate change. If this is achievable in practice, then GN is not unfair.

4.3.3. Privacy and Green Nudging:

The use of GN poses a risk to privacy because informational data about our choices, behaviour patterns, and backgrounds is accessible to the nudgers. A trade-off needs to be made: If informational data are being used to improve behavioural outcomes and seek

collectively desirable solutions, do we value data about our behavioural patterns and preferences more than reducing environmentally mediated harms due to our behaviour?

Such a trade-off also needs to consider the possibility that informational data about nudgee's behaviour can be leveraged to develop long-term habits and attitude shifts required to mitigate climate change. If our informational data can be used efficiently to drive behavioural change and is handled by nudgers who are trustworthy, then privacy concerns associated with GN can be subdued. This proposition is developed in Chapter 6.

However, there is an exception. There is a risk to privacy if informational data are used otherwise, such as sold to third parties or stolen (Ranchordas, 2020). It is a practical concern if our information is used by untrustworthy parties for purposes other than environmental benefits. This point is revisited in Chapter 7.

4.3.4. Autonomy and Green Nudging

As already pointed out, autonomy as 'freedom of choice', as proposed by Thaler and Sunstein (2008), is disputable for GN. Holding onto this understanding of autonomy will not take us far in addressing climate change. If we continue to preserve one's freedom to choose, behavioural shifts towards a healthy planet will be unlikely. However, this does not mean that autonomy is not a consideration for GN to be ethically justified.

Two points about GN and autonomy are notable in this regard. First, whether nudges challenge autonomy or not depends on what we understand by individual autonomy in the context of climate change[74] (Silva, 2022; Siipi & Koi, 2020; Schubert, 2017b). In the context

[74] Bovens (2009) presents two major lines of criticism regarding nudges being in violation of autonomy. First, there is "something less than fully autonomous about the patterns of decision-making that *Nudge* taps into" (Bovens, 2009, 04). Second, when a nudge influences our preferences, we do not know if that influence is a long-term, stable influence, thereby bringing a genuine change in our preferences. By "genuine change", Bovens (2009) seems to imply a coherence between the thought, the reason, and the action of the nudgee. Bovens' (2009) first criticism does not take into account the nature of one's autonomy as it stands concerning the autonomy of others, so it does not contribute to an account of GN. Similarly, the second criticism does not allude to this context because the understanding of 'GN' in this context considers the possibility of influencing deeper cognitive systems. This suggests that nudgees form a coherent set of preferences, develop a moral

of climate change, we need to understand one's autonomy as related to the autonomy of others. To this end, Principled Autonomy, which is O'Neill's (2002c, 2002d) version of autonomy, is more appropriate.

The modern understanding of autonomy, which views autonomy as an individual having complete control over their life and choices, is worthy of some criticism (O'Neill, 2002d). In the context of environmental decisions, O'Neill writes, "If the literature on rational choice has shown anything, they have shown the disjunction between individual choice and collective choice is deeper than once it seemed." (2002c, p.46). She argues for a view of autonomy that recognises individuals as interdependent, and being autonomous means acting within certain obligations or frameworks that make that autonomy and the choices made meaningful (O'Neill, 2002c). From this view, autonomy is understood as acting on principles that can be universal laws. Thus, broadly speaking, for a society to be autonomous, its members must fulfil their obligations and expect others to do the same[75].

Principled autonomy requires the nudgees to fulfil their obligations, thereby respecting the autonomy of others, and acting autonomously as a result. Recall that the individual obligation to address climate change was outlined in Chapter 1: Our individual obligation requires us to be aware of and act with consideration about our behaviour as a link in a series of actions that impose climate-harm. If GN can help us recognise this obligation and act accordingly, then they do not compromise principled autonomy, but help us act in accordance with it. Trust can play a role in upholding such kind of autonomy. This is taken up in Chapters 5 and 6.

---

willingness to act in an eco-friendly way, and learn to reason about environmentally succinct choices, thereby indicating the possibility of a 'genuine change'.

[75] This conception of autonomy is in contrast to individual autonomy, which modern philosophers misinterpret in the term "self-legislation" (O'Neill, 2002b). Self-legislation is not concerned with making one's choices and decisions by oneself. By referring to self-legislation, O'Neill (2002b) emphasises that self-legislation means the legislation (or principles) that can be principles for all.

The second notable point is that autonomy is compromised when one conforms to the expectations of the nudger (Gumbert, 2019b). This means that the nudgee does not act out of sufficient information, cognitive awareness, or moral concerns (Gumbert, 2019b). This objection can be addressed by using nudges beyond System 1 processes to improve nudgee's ability to practice their autonomy. This is to say that instead of merely changing behaviour, we can use GN to influence the existing environmental preferences of the nudgee, not only by providing them with knowledge about the right behaviour, but also by developing the skill-set and moral virtues required in an eco-friendly actor.

4.4. Conclusion:

This chapter reviewed the existing ethical literature surrounding green nudges. There is an urgent need for transitions towards eco-friendly behaviour. For this reason, it was suggested that some of the ethical issues can either be traded off in light of climate change, reformulated, or addressed through an account of institutional trust.

So far, the thesis has responded to two gaps about GN. First, GN is effective with human moral psychology, and therefore, plays a valuable role in addressing climate change. Second, despite their efficacy and promise, green nudges pose ethical burdens. Responding to the two gaps presents an interesting situation–how to balance the promise of GN against the ethical risks imposed by it. To make sure GN has more benefits than potential harms, they ought to be justified from a consequentialist position. To do so, Chapter 5 uses the theoretical apparatus developed in Chapter 2, Environmental System Consequentialism (ESC), to explore why and under which conditions GN can be ethically justified.

Chapter 5. Green Nudging and Environmental System Consequentialism

Chapter 3 demonstrated how GNs are effective in helping nudgees fulfil their moral obligations and act in ways that maximise ecosystemic well-being, despite the evolutionary psychological traits. Chapter 4 then showed how, despite their effectiveness, GN compromises certain objective human values, such as autonomy and fairness, through manipulatively changing behaviours or posing privacy threats. This illustration of effectiveness and countervailing burdens places us in a position to examine whether and under what conditions GN is ethically justified from a consequentialist perspective. The theoretical apparatus developed in Chapter 2, Environmental System Consequentialism (ESC), will be used to do so.

The maxim from ESC prescribes that: An action is justified if it generally maximises ecosystemic well-being, while upholding certain objective systemic values. So for GN to be justified from this position, they ought to be in accordance with the rules that can generally maximise the good outcomes for the ecosystems while upholding the objective values that are (a) important for ecosystems and (b) maintain consistency with other vital human systems. In other words, rather than evaluating whether an action, in this case, the use of GN, leads to good outcomes for an individual, state, or reduces one-time harm, this position requires us to assess GN based on rules and objective values. These rules and objective values are fundamental for ensuring the long-term flourishing of the environmental systems through GN.

Broadly, this chapter frames GN (and other behavioural interventions) in a new light: The cognitive shift associated with GN is justified if the outcome maintains/realigns our interconnectedness with the ecological systems–a reality our current 'autonomous' psychological systems misapprehend. It is worth noting that this use of the term GN moves beyond the conventional literature on nudging, which typically focuses on System 1

processes and choice architecture, to include the broader sense employed by Kaaronen (2018), which encompasses interventions that engage both automatic and reflective cognitive processes. This cognitive shift is necessary. Moreover, such outcomes are not in violation of principled autonomy (introduced in the previous chapter), but are required to uphold it in relation to ecosystemic flourishing.

This chapter presents the argument in the following order. The next section (5.1) reframes the role of GN in light of ESC, summarising its main functions. Following this, (5.2) illustrates what the use of GN, if it follows the principles laid out by ESC, entails for ecosystems and human moral systems. Subsection (5.3) then outlines the rules that GN generally follow or ought to follow to be justified according to ESC. Subsection (5.4) considers the objective ecosystemic goods and the objective goods that need to be upheld to maintain consistency with (or be acceptable within) other systems.

5.1. Reframing the Role of Green Nudging:

As we have seen, from the ESC position, the problem of global climate change is driven through interconnected systems that not only contribute to, but are also threatened by, this crisis. Human psychology and actions underlying these systems have not evolved to deal with this collective crisis; they remain anthropocentric, short-sighted and selfish. Human behaviour, then, pushes the environment and all the systems that are inherently linked to the ecological systems, or more broadly, the Earth system, to the precipice of a catastrophic threat.

Since ESC views the environmental systems as the locus of the moral good, actions are justified if they correct the behaviours, attitudes, and psychological dispositions that mitigate the systemic threat of climate change. It is also morally desirable to consider the possibility of recalibration of human moral psychology, making it equipped to understand

and adequately respond to collective action threats by keeping ecosystems (and not human goods) as a focus.

GN is positioned to drive such corrective mechanisms. GN is justified through ESC not only through correction of discrete behaviours that impose climate-harms, but through three systemic functions: behaviour alignment, formation of virtuous circles, and cognitive shifts.

First, the immediate justification of GN lies in its ability to bring human behaviour into alignment with the conditions of ecosystemic well-being. ESC is concerned with the rules of action. Doing so will involve driving human behaviour in ways that uphold the balance between different *tokens* of the ecosystem and preserving the long-term relationships valuable for the ecosystem to flourish. So from this view, aligning behaviours through GN involves the formation of habits and thus generating patterns of behaviour. GNs alter behaviour patterns like energy-consumption habits, mobility patterns, and eating habits, amongst other things. If these habits and behaviour patterns are adopted by large groups of people, it can lead to a systemic behavioural shift (Nielsen et al., 2024). In this sense, GN can act as a corrective mechanism or tool for behavioural alignment (Kaaronen, 2018).

However, we may take a step further and see whether GN can generate behavioural traits that humans may not yet (arguably) possess. This is the kind of proposal that would fascinate transhumanists. As we do not possess the level of empathy, altruism, and justice to see our actions as morally blameworthy because they compromise ecosysemic well-being (and not just humans in it), a behavioural alignment resulting from the generation of such moral traits would make GN a generative mechanism as well. This generative function of GN may lead to the emergence of a different set of socially acceptable norms, activism to change institutional structures, and a greater level of cooperation.

Although GN can have considerable appeal in the context of climate change, its permissibility as a generative mechanism requires careful consideration. It may allow artificially creating a psychological state of empathy or altruism regarding climate change and other (arguably) moral avenues. Moreover, if being deployed through artificial intelligence systems and their risks, such a generative alignment ought to consider other risks about explainability, trust, and fairness of AI systems.

The conditions of ESC can tackle such risks by appealing to the rules and objective values that are non-negotiable for the justification of an action. So, from such a view, the generation of certain psychological states in human cognitive systems is permissible only if it accords with the rules that maximise good ecosystemic consequences and uphold certain objective values that also intersect with AI systems. For instance, in light of integrating behaviour change with AI systems to generate these states, the objective value of principled autonomy (discussed in the previous chapter and we will see in this chapter) ought not to be compromised. Thus, for behavioural alignment, GN can serve as corrective mechanisms, and possibly as generative mechanisms, as long as the conditions of ESC are met.

Second, the use of GN can create virtuous circles by increasing awareness of one's environmental impact. If habits and behavioural patterns are adopted on a large scale, it will lead to increased awareness of how one's actions influence ecological well-being. Awareness in nudgees will lead to more eco-conscious citizens, economic agents, and political agents. The use of GN and other similar interventions will become less necessary due to increased awareness, and it will lead to the formation of virtuous circles–groups of nudgees who can drive other necessary systemic shifts to address the collective crisis of climate change.

A large number of individuals forming virtuous circles due to increased awareness about their actions may have a profound impact on societies, both political and economic systems. This has the potential to correct human-centric systems, such as better economic and

political systems, in light of climate change. Doing so would mean that nudgees not only become better moral agents, but also better political and economic agents, driving a virtuous change within human-centric systems. Such a systemic change will be desirable for ESC.

However, in light of technological powers being misused by certain actors, objections may be raised against the potential pitfalls of creating such circles. For example, misinformation on fast AI-driven platforms could lead to manipulation and loss of autonomy. Nonetheless, ESC may address this objection by appealing to the objective value of trust (see 5.4.2) that GN ought to uphold in being effective. This means that creating fear or using deception in the creation of such circles would compromise a morally-justified account of trust that is ethical requirement for GN. This account of trust is outlined in the next chapter. Moreover, an effective example of creating such virtuous circles driven through increased awareness is exemplified in the proposal of Behavioural Self-Governance in Chapter 7.

Third, the most profound change possible through GN, as predicted by Kaaronen (2018), will be a shift in the cognitive systems of human beings. Reshaping cognitive systems would involve making nudgees equipped to perceive how indirect effects of their actions trickle through the ecosystem. This involves a long-term shift in the way people perceive their relationship with the environment, looking at their actions as processes within a bigger, interconnected complex ecosystem. This will be desirable from ESC because it can fundamentally change how human systems stand in relation to the ecosystems, creating a balance between human interests and conditions for ecosystemic flourishing. This can be a shift towards ecocentrism that environmental philosophers have long argued for and may prepare us to deal with climate change and other issues of our time.

A proposal of cognitive shifting or shifting of mental models, as Kaaronen (2018) calls them, must also look into the large-scale implementation of such cognitive shifts. This may mean developing a collective intelligence/virtuous circle that can perceive indirect

effects of climate change, and being able to reliably predict future events or catastrophes in this context. This may happen, especially in light of technological progress. However, despite the moral desirability of such a proposal, from ESC, we are compelled to consider the preservation of objective values that are necessary for other human-centric systems to flourish. So, from this view, the cognitive shift may be permissible only to the extent that the objective values of prudence, autonomy, and fairness may be upheld.

Cognitive shifting is a common occurrence–we learn new skills, arts, and often learn to navigate new technologies. This shifting, however, is ethical only to the extent that it does not compromise the cognitive capabilities of an individual to prudently choose to undergo such a shift. Similarly, the nudgee being autonomous in engaging with this change and fairness are important. In the absence of such willingness from the nudgee, such a cognitive shift cannot be a rule that generally maximises the good outcomes for the ecosystem. Thus, from this perspective, a cognitive shift is desirable from ESC only to the extent that it can be a rule and uphold some intrinsic objective values.

5.2. Green Nudging and Systems:

GN enables nudgees to engage in pro-environmental behaviour, improving the behavioural outcomes in a way that is beneficial for the ecological systems. Repeated behaviours can be promising in improving behavioural patterns and can lead to the formation of habits. Developing habits[76] as a result of being nudged depends on how well nudgees understand and respond to the climate crisis. This depends on combining nudging with the strategies of deliberation, education, and participation, an observation not gone unnoticed in the literature on nudging (DellaValle & Sareen, 2020; Lenzi, 2019; Button, 2018). Moreover,

[76] Using GN in this way may also have interesting implications as a method for improving value judgements when examined through Dewey's lens of moral philosophy.

someone who develops a pattern of eco-friendly behaviour, as outlined in (1.5.3), or habits, means that they fulfil their individual moral obligations.

Broadly speaking, this has implications for altering an individual's moral psychology. On an individual level, changes in behaviours leading up to patterns and developing habits in the nudgees are possible when one is aware of their environmental impact. Awareness about one's ecological impact is a signifier of someone who fulfils their obligation to maintain the well-being of the ecological systems, but also has a transformed moral psychology.

While a change in individual moral systems is interesting, a shift in collective moral systems can also be explored through behavioural change and GN. Just like our actions have far-reaching consequences as a result of interconnectedness and interrelationships shared between the components (or *tokens*) of the ecosystems, one's own behaviour is not isolated or disconnected from how others behave/act. This insight is explored in recent behavioural psychology.

Behavioural experts[77] now extensively focus on identifying effective behaviour change strategies by tapping into the mechanisms of how people understand their behaviour in social settings (Nielsen, Clayton, et al., 2021; Nielsen et al., 2024; Nielsen, Nicholas, et al., 2021). This means that changes in behaviour on an individual scale by certain groups/people can signal to others to adapt more pro-environmental behaviours. A collective adoption of such attitudes holds potential for a collective shift in human moral systems to address the climate crisis.

This shift in human moral systems offers an interesting insight for environmental philosophers and similar thinkers. Aside from the use and potential of GN, a change in human moral systems in this context suggests that a shift away from anthropocentrism is

[77] Kristian S. Nielsen's research on behaviour change in mitigating climate change and conserving biodiversity through environmental psychology was found to be notable in this regard. His experiments and insights (published with other experts) on changing behaviour may have the potential to drive collective behaviour shifts through individual behaviour changes.

indeed possible for humans. This is also the kind of expansion of the moral circle that Singer (1981) talks about. Our moral systems and, therefore, our moral judgements are limited in scope (due to factors outlined in Chapter 3). Dealing with the urgent problems of our age, such as climate change, poverty, and animal cruelty, requires us to acknowledge a more non-anthropocentric (or ecocentric, as this dissertation argues) worldview to expand our moral circle. A moral disposition focusing on systems in making moral judgements can help us embrace such a worldview.

Focusing on systems as a moral good, in the case of climate change, means that we will understand how moral judgements are more than a calculation of maximising utility or satisfying preferences. This approach is not only ethically justified but also feasible considering the scientific facts about the systemic nature of the climate issue. Broadly speaking, this motivates the idea that making informed moral judgements means acknowledging the complexity, interdependencies, and relatedness between various stakeholders who are concerned with the effects of our actions. This dissertation aims to highlight this aspect of morality by focusing on the nature of systemic goods, particularly their significance in the context of climate change.

In the context of GN, establishing rules about seeking ecosystemic goods is crucial for developing long-term behavioural patterns. Moreover, making rules about the ethical use of GN can generalise 'ethical nudging practices' which can serve as a template for the design, implementation and regulation of GN strategies. This will be an important insight for nudgers, as the ethicality of GN hinges on how nudgers understand, use, and implement this strategy.

Upholding certain objective goods through GN is also critical in this context. On a practical scale, such a promise of GN will not be possible without combining it with technological and scientific advancements. This means that behaviour change towards pro-

environmental outcomes will be possible through carbon footprint calculators, mobile applications, and Internet of Things (IoT) systems.

This, as Chapter 4 has outlined, raises ethical concerns about human autonomy, fairness, and human prudence (or the ability to make well-informed good decisions on one's own). This suggests that the conditions under which GN can be ethically permissible will require consideration of these aspects in relation to other important human values. Thus, it is both timely and practical to outline these objective goods/values.

From the position of ESC developed in this thesis, GN may be justified depending on whether it maximises the good outcomes for ecosystems through its influence on human behaviour, patterns of behaviour, habits, and moral shifts. Its ethical permissibility, as we will see in the following subsections, also depends on whether it preserves/upholds certain objective goods for the ecosystems and the other existing human systems.

The next two subsections justify GN from ESC in the following order. Subsection (5.3) outlines how GN leads to /shows promise in leading to actions that are in accordance with the rules that maximise good ecosystemic outcomes. Following this, Subsection (5.4) considers the objective goods that GN ought to uphold to be ethically permissible.

5.3. When are Green Nudges Ethically Justifiable?: The Rules

The subsection outlines three ways in which GN helps nudgees act according to the rules that maximise good ecosystemic outcomes. Overall, GN leads to good ecosystemic outcomes by improving behaviour, increasing one’s awareness about their behaviour and its relationship to climate change (which leads to virtuous circles), and reconfiguring our cognitive systems to better perceive our relationship and role in contributing to climate-related harms.

Additionally, in a recent piece, DesRoches et al. (2026) recommend a ‘Variable Nudging Model’, which considers nudging as a matter of degree, rather than as a simple

binary choice (p.38). This resonates with the idea that GN needs to be calibrated and adjusted for the context, reflecting the conditions for ethically justifiable GN through ESC: behavioural alignment, virtuous circles, and cognitive shifting.

5.3.1. Behavioural Alignment

Nudges are successfully applied by government units and by private and public actors to promote a broad range of sustainable behavioural outcomes (Sunstein & Reisch, 2014). Since human ends are morally justified to the extent that they promote systemic well-being, GN directs nudgees toward behaviours that align with ecosystemic outcomes. In doing so, GN follows the rules of ESC.

The first reason why GN improves systemic outcomes is that it addresses the harm inflicted by a nudgee's actions on other *tokens* in the ecosystem. As emphasised in Chapter 4, GN addresses (to some degree) third-party harm to other humans. For a policymaker, the justification for using GN to influence a nudgee's choice flows from their duty to optimise pro-environmental outcomes. Therefore, GN that (1) prevents the nudgee from inflicting harm on others due to their behaviour, (2) reduces third-party spillover effects, (3) increases social welfare without compromising the well-being of the nudgee is permissible according to the 'harm principle' ingrained in liberal theories (DesRoches et al., 2023). Another condition can be added to the third-party justification of GN offered by DesRoches et al. (2023). GN also prevents third-party harm by (4) preserving the relationships between various components of the ecosystems, which are critical for protecting overall ecosystemic well-being.

Adding this factor in evaluating third-party harm is crucial because embedded anthropocentrism in human psychology has overlooked the long-term value of other components of the ecological systems (Singer, 2011). Our actions are primarily directed towards improving human-centric outcomes. The moral imperative to address climate change requires us to shift the focus from human benefits to environmental outcomes. Thus, in addressing the third-party harm, GN not only optimises well-being from a human-centric perspective, but also from an ecocentric point of view. For instance, our current ends, through

nudges, such as reducing greenhouse gas emissions, frugality in lifestyle, reuse and recycling, can have future implications in preserving biodiversity (and other tokens) on the planet.

Second, GN leads to the formation of eco-friendly habits, ultimately developing eco-conscious citizens who contribute to environmental goodness over the long term (Hukkinen, 2016; Kaaronen, 2018). This can be a rule for maximising good outcomes in the context of climate change. Certain nudges send prompts to the user at specific times of the day. Such prompts range from reminders to turning off lights and closing windows during heating times, which can effectively reduce energy consumption (Bonini et al., 2017). When such well-integrated nudges within one's lifestyle can result in the formation of eco-friendly habits, they also contribute to ecosystemic well-being.

Using default choices to make nudgees accustomed to specific options can also contribute to habit formation. For instance, when choosing energy operators, placing green energy as a default option increased the sign-up for renewable energy in the Black Forest region in Germany (Pichert & Katsikopoulos, 2011). Similarly, double-sided default printing saved over five hundred trees in a single semester at Rutgers University (Schubert, 2017b). Such nudges habituate one to engage in an eco-friendly behaviour, ensuring that an individual upholds the well-being of environmental systems even without a nudge. Moreover, habit formation is also desirable from a virtue-ethics perspective, which Hiller (2013) acknowledges can align with the theory of system consequentialism.

Third, until the habits are developed and patterns of behaviour emerge, GNs can reduce the cognitive effort required to engage in pro-environmental behaviour.[78] This can lead to immediate outcomes. This is possible by communicating information about eco-

[78] It can be debatable whether this can be a rule that generally maximises pro-environmental outcomes. An interesting debate in this context is offered by Lily Eva Frank (2020), who explores whether moral struggle is a requirement for moral progress. Nonetheless, making it easier to engage in eco-friendly behaviour can be a temporary rule, as it leads to immediate positive outcomes for the ecosystems (which is urgent considering the climate crisis).

friendly alternatives or by using default options. For instance, environmentally related information is often presented to people without taking into account their cognitive processing capabilities at a given point in time (Lehner et al., 2016). GN can restructure the way environmentally relevant information is presented to the nudgees. For instance, Kallbekken and Sælen (2013) show how changing the decision-making environments, such as changing dish shapes or encouraging posters to take multiple servings of food to avoid food waste, significantly changes the behaviour of buffet guests.

Similarly, using 'Desktop Orbs', a ball that changes colours based on the energy efficiency of households, shows promise in signalling consumers when to reduce energy consumption (Selinger & Whyte, 2011; Hausman & Welch, 2010). Default options are the automatic options pre-selected by the nudger for the nudgee. In the absence of an active choice, a default option can help nudgee people to make the right decisions. Some examples of default options, such as signing up for renewable energy or vegetarian menus, can make this possible (Pichert & Katsikopoulos, 2012; Schubert, 2017b). These examples illustrate how GN can make it easier for nudgees to engage in pro-environmental behaviour.

5.3.2. Awareness

Apart from achieving behavioural outcomes, GN can make us much more aware of how our actions have an environmental impact. These actions can be rules about promoting eco-friendly behaviour. GN can do so by appealing to social norms and self-image for moral motivation, and providing cognitive assistance when one lacks information about the right environmental choices. Increased awareness can lead to long-term change in environmental attitudes, habits, and behaviours, forming virtuous circles of nudgees who are capable of driving systemic shifts necessary to address climate change.

GN can influence the social norms surrounding pro-environmental behaviours. Recall from Chapter 3 that humans desire a positive social image, which is an evolutionary feature

(Tajfel & Turner, 1985). To generate awareness about climate change, it is possible to appeal to social image and shape the social norms surrounding pro-environmental behaviour. A previously discussed example of this is Nagatsu's (2015) study on the "Don't mess with Texas" campaign, which significantly reduced littering behaviour (Footnote 42).

Li et al. (2024) show that social media platforms offer a promising avenue for shaping social norms related to pro-environmental behaviours. These norms can be understood as the acceptance and expression of shared values that emerge from collective knowledge about what people ought to do in response to the climate crisis (Li et al., 2024). Another way to frame social norms is by linking GN and similar behaviour change strategies with models of self-governance, proposed in Chapter 7.

An effective way to shape social norms through GN is by leveraging social media connectivity. An example of this is PeakApp[79], an energy conservation program, which aims to drive long-term energy-saving behaviour in households by nudging users to maintain consistent energy habits through continuous engagement on social media platforms.

Social media platforms accelerate norm shaping in three ways: incentivising moral outrage, allowing context collapse, and increased social learning (Brady & Crockett, 2023). Moral outrage is amplified because social media platforms benefit from increased user engagement. As a result, content with moral and emotional appeal is more frequently shared and interacted with. Such content leads to "context collapse", where imaginary groups emerge irrespective of their individual identities. Additionally, in such groups, the speed of normative signalling is higher, leading to increased social learning as the ideas of right and wrong are quickly acknowledged on social media platforms (Brady & Crockett, 2023). The example of PeakApp suggests that these mechanisms may absorb pro-environmental

[79] PeakApp is community-based energy conservation program that aims to serve the EU's target for smart-meter roll out to transition energy efficient citizens consumers. https://www.peakapp.eu/project-overview/

behaviours as social norms. While ethical concerns around user manipulation on digital platforms remain, combining the potential of social media with behavioural change through GN can have long-term implications in shaping social norms.

Another way to increase awareness about pro-environmental behaviour is by appealing to the self-image of the nudgee. A positive self-image is psychologically desirable (Schubert, 2017b). By associating self-image with good environmental outcomes, GN improves awareness of one's environmental behaviour. One way to achieve this is by increasing the information salience. For instance, a field study by Bolderdijk et al. (2013a) shows that appealing to a positive self-image as an environmentally conscious citizen leads to improved behaviour in personal transport choices. Prompting users about environmental concerns to get regular tyre pressure checks was more effective than appealing to financial savings.

However, such appeals to self-image through information salience are observed to be effective only to the extent that the nudgee is driven by personal values. In the absence of environmental concerns, GN via self-image provides mixed evidence (Bolderdijk et al., 2013b). Thus, it indicates the need for a cognitive shift to internalise environmental values in the nudgee for better ecosystemic outcomes.

5.3.3. Cognitive Shifts

Recall that Chapter 3 argued that GN helps us act despite the psychological traits that limit our ability to make well-informed and ethically sound decisions in the context of climate change. Beyond behaviour change and awareness, there is also a potential for GN to influence our understanding of the ecological crisis and our relationship with nature (Kaaronen, 2018). Rules about such a cognitive shift, if possible, can have long-term positive outcomes for the human-nature relationship and the well-being of the ecosystems. This would be the most profound shift possible through GN.

First, it is possible to perceive human actions as processes through GN. The ecological crisis results from complex and interconnected processes. However, our cognitive systems do not yet understand climate change[80] as a slow-moving process (Hukkinen, 2016). For instance, landslides result from a series of environmental processes. Extreme rainfall, caused by temperature changes, washes the soil into the rivers. As a result, the rocks, trees, and other debris that were held together by the mountain soil get destabilised. This material is pulled down the slopes of the mountain due to gravity. It results in landslides, which bring about serious consequences. To see outcomes in such a way is to recognise that all consequences result from processes constituting interactions between different *tokens* of nature (Kaarnoen, 2018).

Human cognition is not yet equipped to perceive climate change in this way. To address this limitation, we need a shift in our cognitive systems so that we perceive our interconnectedness in the environmental processes and understand our contribution to climate change.

GN can achieve this shift. Kaaronen (2018) focuses on the role of GN in developing a new environmental consciousness. They argue that it is possible to nudge the mental models of nudgees (or, in other words, achieve the cognitive shift) by helping them understand the interconnectedness between themselves and nature, and how their actions contribute to the processes that result in harm (Kaaronen, 2018).

GN can show us how our ends are not separate from the ecosystem, but they are a part of it and contribute to its well-being. It is possible to do so by emphasising how our actions are taking place in a bigger system. For example, Hilton et al. (2018) suggest that GN can

[80] Despite having a long history of thought that humans experience the world through changes, dating from pre-Socratic philosophers like Heraclitus, almost all human endeavours have focused on substantive or measurable outcomes. This is a substance-based approach embedded in our mental models (Kaaronen, 2018). A substance-based approach also underlies the way we perceive climate change.

reduce electricity consumption by showing images of polar bears on a floe, where the health of polar bears and the size of the floe depend on one's energy behaviour. By doing so, GN can tap into moral traits like altruism and empathy. It extends our moral concerns towards other beings in the ecosystem.

Another way to do so is by emphasising the cumulative effects of one's actions on the ecosystems. Recall the example of taking a drive in a gas-guzzling SUV. GN, through GPS trackers, can show how the combined effects of people speeding SUVs for a Sunday drive impact rainforests.

Moreover, GN can also appeal to the future consequences of our actions. For example, smart energy-saving applications can monitor electricity usage and project the long-term environmental impact of one's consumption patterns. Such future-related visualisations framed in an accessible manner can invite the nudgee to adopt habits that reframe the way we perceive climate change, and our relationship with it.

Second, while this is an instance of changes on an individual level, a large number of people experiencing such cognitive shifts can be transformative. When we start looking at our actions as processes, it reframes our decision-making and leads to the cultivation of certain habits. Such habits may involve a forethought about the environmental consequences of one's actions, an increased sense of justice towards those suffering due to the effects of climate change, and a sense of empathy for other human and non-human beings. When cultivated in large numbers of people, such habits can drive large changes in economic and political systems.

Thus, the possibility of achieving large-scale cognitive shifts can mark a moral transformation–something that has long been desired by philosophers, environmentalists, and visionaries. For instance, a cognitive shift towards understanding our actions as participatory in climate harm can develop 'respect for nature' that Taylor (1981) argued. By making us

equipped to see ourselves as part of the ecosystemic processes, GN can help us achieve 'oneness' with the environment that Naess (2005) desires. Similarly, for Singer (1975), this would be imperative because it will preserve the well-being of sentient creatures, a moral requirement in his theory of sentientism. Lastly, from a system consequentialist position, such cognitive shifts can result in lasting ecosystemic outcomes that preserve its conditions for flourishing.

Thirdly, this cognitive shift also means that GN can not only help us achieve our moral obligations, but also make us better equipped to fulfil these obligations[81] on our own. Cognitive shifts can be made possible by making these processes psychologically accessible to us. As a result, they provide an opportunity to internalise our obligations.

When we act to maximise the ecosystemic wellbeing, we also act in our own interest, as we are not separate from the best state of affairs for the systems. GN can frame our actions as part of broader ecological processes. When this framing enables us to act, we fulfil our moral obligations on an individual level. Moreover, once this cognitive shift[82] occurs, we become more accustomed to making decisions that preserve systemic goodness, even in the absence of nudging.

ESC also holds that in acting according to rules, some objective values ought to be upheld. This is important because these objective values/states/goods are intrinsic to the long-term well-being of other valuable systems (that are also connected to the ecosystems). These objective values can be seen on two levels. First, the ecosystemic objective values that are

[81] Recall from Chapter 1 that we have a collective obligation to address climate change. There are various reasons why this obligation exists. As climate change affects ecosystems, and our actions contribute to it, we ought not act in ways that harm the ecosystem. A justification for such an obligation results from the idea of systemic goodness, which is argued as the supreme consequentialist good. This obligation also applies on an individual level, as outlined in Chapter 1. Because our actions occur within systems and affect ecosystemic well-being, the moral obligation is both collective and individual.

[82] This shift is a process as well, which will evolve as we become better environmental actors.

necessary for the ecosystem/environmental systems to flourish. Second, values intrinsic to human-centric systems–economic and political systems. If GN is to be accepted as a legitimate and pragmatic strategy for achieving ecosystemic goods, certain objective values, like trust, principled autonomy, fairness, and prudence, must be defended.

5.4. Objective Values that Green Nudging ought to Uphold:

While GN improves systemic outcomes, certain objective values should not be compromised, as these values achieve long-term systemic benefits and prevent spillover effects to other systems. This means that, as ecological systems are interconnected with other systems, such as economic, social, and political systems (as discussed in Chapter 1), upholding some objective values enables us to maximise the ecosystemic consequences while ensuring the functioning of these other systems. The following section outlines systemic objective values, including ecosystemic objective values and those that are universal and necessary for any system to flourish.

5.4.1. Ecosystemic Objective Goods

This dissertation has focused on upholding ecosystemic good as the consequentialist value. Ecocentric philosophers, such as Aldo Leopold (1949), Holmes Rolston III (1975), and Arne Næss (2005), have outlined, to a great extent, the conditions and state required for the goodness of these systems.

In the context of climate change and GN (and behaviour change), a position informed by ESC ought to consider the ecological health and well-being. Broadly construed, we are concerned with the objective values that can ensure good outcomes for ecosystems. Three objective goods /values[83] are outlined in this context.

[83] In using the terms 'goods' and 'values' interchangeably, I am essentially implying the objective values that generate states/outcomes that are desirable by ESC, and objective goods that are necessary for these objective values or outcomes to be realised. For instance, the discussions here do not explicitly outline the objective value of 'ecological integrity', a theme that recurs in Leopold's (1949) scholarship. Nonetheless, the resulting

First, ecological flourishing is necessary for ecological health. To flourish is a non-negotiable component of a good life. Our ethical sequence, as Leopold (1949) argues, has focused on the relationships between humans. Extending this ethical sequence to ecological systems means that we ought to consider ecological flourishing as a morally desirable state. Just as ethical judgement involves ensuring what is good for other individuals concerned with an action, in the case of ESC, a judgement requires considering what makes ecosystemic flourishing possible.

So in considering human actions possible through GN, one of the main objective goods is ensuring that GN actually contributes to actions that make ecological flourishing possible.

Second, as system consequentialism prescribes, we ought to consider maintaining the relationships and networks that are necessary for systemic goods. Chapter 2 has argued for such an understanding of a system, where different *tokens* and the relationships between them comprise the system. Thus, such relationships are important for ecosystemic flourishing. GN can encourage actions that sustain and improve these relationships.

Third, as ecosystems are vital to the Earth system's well-being and existence, maintaining ecosystemic balance is critical. As climate change is an issue that generates vulnerability, suffering, and adverse consequences across various components of the planet, it disrupts the balance of ecological systems. This means that the outcomes of GN should be such that it preserves the long-term balance of the Earth's ecosystems.

objective goods are implied in this context. This is to say that ecological balance or flourishing is not possible without upholding ecological integrity.

5.4.2. Objective Values for Acceptability in Human-centric Systems

The following section considers the objective values that GN ought to uphold to ensure that existing human-made systems, which are nonetheless critical for the well-being of humans (who are components of ecosystems), are not disrupted.

Prudence. GN can improve ecosystemic outcomes if it fosters prudence in the nudgees. Prudence involves possessing foresight, deliberating about one's environmental impact, and taking into account the effects of one's actions on the future. When nudgees are prudent, they contribute to long-term behavioural shifts that improve ecosystemic outcomes even in the absence of nudges.

For ecosystemic flourishing, nudgees should internalise their obligations and continue to engage in eco-friendly behaviour. This can lead to a large-scale cognitive shift where environmental concerns are reflected on a behavioural level, but are also manifested on a collective scale to drive systemic changes, such as social and political changes, to better address climate-harms.

Some studies provide evidence of such potential in GN. For instance, nudging for charitable giving enhances altruistic tendencies in nudgees that last beyond the presence of a nudge (Capraro et al., 2019). Similarly, combining nudging with carbon taxes increases the salience of taxes, exemplifies pro-environmental concerns in public, and improves public acceptance of carbon taxes (Gravert & Shreedhar, 2022). This suggests GN can foster prudence in nudgees regarding their environmental behaviour.

However, nudging may have adverse effects in this context. For instance, the implications of using GN is negative when people are left with a fragmented self, which can leave them unclear about their environmental preferences (Bovens, 2009).

To avoid such cases, nudging can be combined with effective framing of narratives around climate change by local governments and the media, mostly based on local values and norms (Romsdahl et al., 2018). Lenzi (2019) also notes that, "Combining deliberation with such nudging strategies promises to be timelier than relying on deliberation alone and can be done in democratically legitimate ways." (p 326). In other words, reforming media systems, strategising communications, framing debates, and incentivising pro-environmental behaviour are some of the strategies that can be combined with nudges (Lenzi, 2019). Doing so can minimise the risk of distortion of the preferences of nudgees. To further develop this proposal, Chapter 7 explores whether GN should be combined with participatory governance, and can preserve and develop the quality of prudence in the nudgees.

Fairness. Moreover, to consistently uphold ecosystemic goodness, GN should be fair in distributing the benefits and burdens among nudgees. A similar argument suggests that the urgency of climate action does not justify imposing disproportionate burdens on disadvantaged groups (Steel et al., 2025).

Recall that the cost-benefit assessment of GN takes into consideration whether or not the benefits of GN to third parties outweigh any harm imposed on the nudgee (DesRoches et al., 2023). To be fair, nudges should promote behaviour that does not disproportionately target or make certain groups of people vulnerable.

Suppose the design and implementation of GN is not perceived as fair. In that case, GN risks spill-over effects, such as reduced voluntary public support for behavioural change, increased economic and social vulnerability, and loss of autonomy. These effects harm nudgees directly, but also compromise the goal of ecosystemic well-being.

ESC considers the rules that can maximise long-term good systemic well-being. Rules that ensure nudging is fair to the nudgees, secure the conditions for the ecosystem to flourish, and thereby maximise good outcomes for a system. This is important because the

ecosystemic well-being depends on the nudgee as well. When nudgees are lied to, manipulated, or exploited, it also impacts the system.

There can be several ways to argue that the well-being of the nudgee is required from ESC. The first way is to argue that nudgees are *tokens* of the system. Any harm to the nudgee disrupts the systemic goodness. The second way is that, as nudging is a non-coercive intervention, unlike taxation or mandates, it is important for the nudgee to be convinced by the goal of behavioural change. In the absence of fair nudging processes, the nudgees can refuse to cooperate. The third way is to appeal to the *better/worse system* premise by Hiller (2011, 2013). There can be two possible worlds, one in which GN is fair and leads to pro-environmental behavioural shifts, and the other in which GN is unfair but still leads to equally good environmental outcomes. The system where nudgees are treated fairly is still better than the other one because the nudgees are valuable parts of the ecosystems. Hence, being fair is important while pursuing good systemic outcomes.

This logic can be applied to all the objective values discussed above. A system with prudent nudgees, fair nudging procedures, autonomous nudgees and nudgers, and a morally justifiable kind of trust between the nudgees and the nudgers is a better system than one where these values are absent[84].

To ensure fair GN practices, three concepts of institutional justice can play a crucial role: distributive justice, procedural justice, and recognitive justice. Distributive justice requires fair distribution of benefits, burdens, opportunities, and responsibilities (Miller, 2021). Procedural justice demands that all participants affected by an environmental outcome have a role in shaping the strategies that lead to those outcomes (Sustainability Directory, 2025). Recognitive justice involves being adequately and appropriately recognised as

[84] There can be rare instances where these objective values clash with systemic goods. In such cases, we should still preserve objective values as there can be epistemic limitations of our foresight about the valuable systemic outcomes.This reveals a potential drawback of a system consequentialist position.

possessing a moral value (Miller, 2021). A proposal of behavioural self-governance and trust between the nudgee and the nudger provides a way to maintain fairness in the pursuit of ecosystemic goodness. Chapter 7 takes this into account.

Principled Autonomy. Principled autonomy, as outlined in Chapter 4 and based on O'Neill's (2002b, 2002c, 2002d, 2002e) conception, involves acting in ways that are universalisable. This type of autonomy requires our rights to be protected, such as the right to live in a healthier environment, but also requires us to act in ways that uphold the rights of others to do the same.

Principled autonomy is instrumental for upholding ecosystemic well-being. When our actions cause harm to third parties, we violate the autonomy of those who deserve the right to live and flourish within healthier ecosystems. O'Neill (2002a, 2002b) argues that it is not possible for one to have their rights upheld when others do not protect them. As long as humans fail to reduce their carbon footprint, the possibilities for the ecosystems to exist and flourish will be violated, and the well-being and rights of the beings dependent on those systems will be compromised. Thus, there is a need for nudgee to protect the rights of other systemic *tokens* by changing their behaviour.

When nudgees have reduced their environmental impact, are aware, and possess cognitive capabilities to understand the impact of their environmental actions, they ensure the conditions for other *tokens* in the ecosystems to flourish. This means that in doing so, nudgees act in ways that protect the rights of others affected by their behaviour, thereby satisfying principled autonomy.

However, we must also ask whether the nudger protects the rights of the nudgee. Such rights involve being nudged fairly, truthful feedback about one's behaviour, and being provided with the right knowledge about environmental outcomes and the ways to achieve

them. In other words, the nudger must act in ways that can generate the conditions for the nudgee to truly fulfil their (principled) autonomy. This means that the nudger should fulfil their duties to protect the rights of the nudgees.

As discussed in the review of ethical themes and issues surrounding GN, there remain concerns about manipulation and unfairness. In the context of GN, a nudgee's autonomy cannot be preserved if the nudgers do not fulfil their obligations. One of those obligations is to be trustworthy towards the nudgee, a proposal developed in the next chapter.

Trust. We cannot have rights if others do not fulfil their obligations (O'Neill, 2002a). Principled autonomy requires us to have universal rights, but also requires fulfilling one's obligations so that others can have rights. For instance, we have a right to health because doctors fulfil their obligations as medical practitioners. For patients to have rights, doctors should have their rights[85] upheld. In short, O'Neill (2002a) argues that in demanding rights, we must also ensure that we uphold our obligations towards others.

Trust offers the basis for principled autonomy (O'Neill, 2002d, 2002f). Principled autonomy requires us to act according to principles that can be applied to all. This entails rejecting coercion, fear, and deception, as these cannot be universal principles. For instance, if coercion were allowed by a universal principle, it would allow mistreatment of all beings, posing harm to them, which would ultimately compromise the ecosystemic well-being.

Trust in the right places and to the right extent can avoid the risk of being deceived or coerced. This means that trust can preserve conditions for principled autonomy to flourish. Placing trust carefully means we avoid trusting people who can take advantage of us.

[85]The issue of the safety of the medical community during late working hours in India sparked massive outrage in the state of West Bengal last year (Raj, 2024). It exposed shortcomings of India's health infrastructure, including a lack of security protocols for doctors, political pressure under which doctors operate, inadequate emergency response systems, and insufficient legal protection for healthcare workers. Such vulnerabilities endanger the lives and rights of medical staff and also have long-term adverse consequences for patient rights, trust, and the overall health infrastructure in the country.

However, we need morally justifiable trust. Too much trust or too little trust can destabilise society or increase the chances of being deceived or generate excessive fear, respectively[86]. In what follows, Chapter 6 will explore what constitutes a morally desirable type of trust and what is required to uphold it in the context of GN. The chapter argues that two rules are required to have the right kind of trust: the trustor (the nudgee) assesses trustworthiness, and the trustee (the nudger) demonstrates trustworthiness.

Trust is also a functional value necessary for other objective values, such as prudence and fairness. A nudgee can practice prudence when they are able to assess the trustworthiness of behavioural change strategies suggested by the nudger. Similarly, when a trustee (or nudger) demonstrates trustworthiness, and a trustor (a nudgee) is active in examining the extent of trustworthiness, they can call out if the practice of nudging is unfair. The next chapter justifies such an account of trust and shows how these objective values can be preserved by building trust between the nudger and the nudgee.

5.5. Conclusion:

The theoretical framework developed in Chapter 2, Environmental System Consequentialism (ESC), was applied to GN. This provided a new understanding of what GN can achieve from a systemic perspective: behavioural alignment, virtuous circles, and cognitive shifts. The rules that can ethically justify green nudges were illustrated. The ecosystemic objective values and the objective values required to maintain consistency of green nudges with other vital human systems are also considered.

[86] For instance, if people are too trusting, they will allow others to treat them as they wish. This would lead to a society where deception becomes a universal practice. Without sufficient trust, it would be chaotic, no one would cooperate, and everyone would rebel and resist. Society will break down. This is also a state where GN may not be able to maximise ecosystemic flourishing.

Within the objective values that are necessary for GN to maintain consistency (and thus, acceptability) with human-centric systems, trust is the fundamental objective value that makes it possible for other objective values to exist, rules to be followed, and GN to be effective. In other words, as we will see in the next chapter, trust is the functional criterion that makes it possible for GN to be ethical and effective. In the next chapter, O'Neill's (2018, 2002a) scholarship on trust and trustworthiness provides a basis to outline the main principles for placing trust in an ethically justified manner. Potential signifiers of trustworthiness in the context of nudging are also considered.

Chapter 6. An Account of Trust in Green Nudges

The previous chapter argued that GN is ethically justified from ESC when it follows rules that uphold systemic goodness and preserve certain objective values. Trust is a functional objective value in this context. In other words, rules governing morally justifiable trust are required in GN to maximise ecosystemic goodness.

The goal of this chapter is to explore the role of trust in GN. In doing so, the rules about trust that justify GN from an ESC position are outlined. The chapter also shows how other objective values are upheld by building a morally justified form of trust between the nudgee and the nudger.

To sketch an account of trust in GN, this chapter draws on Baier's (1986) perspective to examine how trust is the basis of moral relationships. However, Baier's account has limits. It lacks applicability to institutional trust. O'Neill's (2002a, 2002f, 2018) account of trust and trustworthiness fills this gap by outlining the moral grounds for placing institutional trust. The chapter illustrates trustworthiness and trust in GN by adapting O'Neill's approach. The features of trustworthiness presented here are preliminary. They can be adapted or modified based on the social, cultural, and economic contexts in which GN operate.

To achieve the objectives set above, the chapter is structured as follows: The next section outlines why trust is important from an ESC position and the applicability of Baier's (1986) and O'Neill's (2002a, 2018) accounts of trust and trustworthiness for nudging (Section 6.1). Subsequently, the nature of the trust relationship between two parties and the rules about trustworthiness in institutional contexts are outlined (Section 6.2). Lastly, this framework is adapted for GN (Section 6.3) through two rules: the nudger demonstrates trustworthiness and the nudgee assesses trustworthiness.

6.1. The Basis for Trust in Green Nudges:

This section argues that institutional trust between the nudgee and the nudger are necessary for GN to contribute to the systemic good. The following two subsections develop this position. First, rules about trust are ubiquitous and are important for upholding principled autonomy (6.1.1 and 6.1.2). Second, institutional trust between the nudgee and the nudger can increase the chances of effective GN. Moreover, building institutional trust increases the extent of perceived fairness and promotes prudence in trustors (6.1.3).

6.1.1 The Need for Trust:

Humans are familiar with the attitude of trust. We exhibit trust towards family members, friends, peers, strangers, and enemies (such as, to follow the ceasefire) as well. Trusting another being is a universal activity, observed in humans and in animals (Whitehouse, 2024). From an evolutionary perspective, an attitude of trust towards predecessors enabled humans to survive (Whitehouse, 2024). Humans have learned survival mechanisms by trusting and conforming to the rules and rituals set by their groups, without relying exclusively on shared reasoning and evidence (Whitehouse, 2024). Noting the ubiquity of trust, Baier (1986) writes, “We inhabit a climate of trust as we inhabit an atmosphere and notice it as we notice air, only when it becomes scarce or polluted.” (p. 234).

Trust is pivotal for a functional society. At least in formal relationships where personal interactions are limited, we must trust others due to our vulnerability to things or events beyond our control (Baier, 1986). We trust doctors to look after our injuries and ailments, teachers to look after the education and safety of our kids when they are in school, and our employers to pay us at the end of the month. One’s vulnerability towards another is a brute fact; trusting others to take care of some matters is inevitable. We trust others when we are vulnerable, aware of our vulnerability, or expect to be vulnerable due to the actions of others that can have implications for us.

There are at least two parties involved in the activity of trusting. The person or party who places trust in another, in the technical jargon, is called the trustor. The person or party in whom one places trust can be called the trustee. The matters for which the trustor engages with the trustee are called entrustment or entrusted matters (Baier, 1986).

Trust is an attitude; it has a psychological element to it. In trusting another person, the trustor places reasonable belief or confidence (i.e., maintains such an attitude) that the trustee is capable of/willing to/committed to taking care of the entrusted matters, or the domains in which one is vulnerable (McLeod, 2023).

Since trust is an attitude, it can be misplaced by the trustor in a manner that exposes them to risk of deception, coercion, and manipulation (Baier, 1986; O'Neill, 2002a). Placing trust in criminals and cheaters is an example of such trust. Such malicious actors understand the vulnerability of trustors and invite trust only to break it to their advantage. The common occurrences of deception on digital platforms, such as Ponzi schemes and fraud, are examples of misplaced trust.

To avoid such dangers, placing trust rightly is important because it lets us count on others, safeguard our entrusted matters, and prevent harm (Baier, 1986). Trust is the foundation of any moral relationship and is essential for a well-functioning society, as it protects one’s vulnerabilities (Baier, 1986).

The role of trust in GN and its impact on systemic well-being is significant due to the functional value of trust. Trust between the nudgee and the nudger is required in GN, just as it is foundational in any moral relationship. For ESC, activities that lead to better ecosystemic outcomes are morally justified, and trust plays a crucial role in achieving such outcomes.

In the context of nudging, a nudgee is the trustor and the nudger is the trustee. The nudgee is a trustor because they are exposed to countervailing risks, such as loss of autonomy

and unfairness. As the nudger is in a position to address such harms, such as by using GN legitimately to improve environmental preferences of the nudgee, incorporating the nudgee in the process of environmental governance, and minimising the scope of deceptive actors to leverage behavioural insights[87], they are in a trustee's position.

Rules for placing trust carefully are essential in this context, as the absence or excess of trust has implications for the outcomes of GN. When trust is misplaced, nudgees may, for instance, interpret misinformation as a scientific insight commonly perpetuated by other market actors (Braganza & Kapeller, 2025). Alternatively, if trust is withheld when needed, nudgees may assume that the social norm information provided by nudgers on energy consumption is patronising or insulting (Allcott, 2011). Or, in another scenario, too much trust can lead one to agree to coercive practices. In all three cases, the outcomes aimed for by GN are compromised.

6.1.2. Trust and principled autonomy:

By trusting according to certain rules that lead to better outcomes, we can minimise the possibility of these practices and thus create the conditions for nudgees to be autonomous[88]. In other words, placing trust in a justified manner can be a universal rule and makes room for principled autonomy to flourish.

Recall that principled autonomy also requires us to fulfil our universalisable obligations. In referring to the principled autonomy in this thesis, two notions of obligations are referred. One that requires us to act in accordance with our individual obligations to

[87] The theme of consumptive nudging is not an ethical concern associated with GN per se, but the use of such practices has implications for the outcomes of GN. A focused discussion of this problem is considered in Chapter 6.

[88] To revisit the argument outlined in the previous chapter: Trust provides conditions for principled autonomy to flourish (O'Neill, 2002d, 2002e, 2002f). Recall that principled autonomy requires acting in ways that can be universalised. Misplacing trust cannot be a universal rule. If they were, the outcomes for society would be catastrophic due to deception and coercion.

address climate change. GN helps us fulfil such individual obligations and make us autonomous.

The other type of obligation requires us to avoid, flag, and resist deception and manipulation in GN. Doing so is desirable for nudgers and nudgees. In the absence of this, deceptive or manipulative GN practices compromise the long-term goodness of the ecosystem and its consistency with other systems of value, for example, a system of democratic governance or a system of the state. These obligations are fulfilled when nudgers and nudgees act according to universalisable rules about placing and inviting trust based on evidence of trustworthiness. In summary, morally justifiable trust is essential for principled autonomy. The next section explores how trust enables practical, fair, and GN practices with prudent nudgees.

6.1.3. Practical Advantages of Trust for Effectiveness, Prudence, and Fairness:

Trust also has practical advantages in terms of effectiveness, fairness in nudging practices, and fostering prudence in nudgees. There is empirical evidence demonstrating the potential of trust between the nudger and the nudgee for effective GN. In examining the success of nudging through eco-labels, Daugbjerg et al. (2014) observed that the extent of trust in the government (i.e., the nudger) was directly proportional to the effectiveness[89] of the labels.

Similarly, a weighted regression analysis strategy demonstrates in high confidence that trust in nudgers is positively correlated with the approval of nudges by the nudgee (Sunstein et al., 2019, p 83). The study shows that citizen perception and judgement of nudges and other behavioural strategies in diverse contexts, such as Ireland, Germany, Mexico, China, and the USA, finds trust as a key factor for the effectiveness of nudging (Sunstein et al., 2019). Thus, this suggests that fostering trust can play a role in achieving behavioural outcomes that are effective in encouraging sustainable behaviours.

Similarly, GN can be unfair in terms of its redistribution effects and tendencies to divert attention from the institutional problems underlying climate change (Schubert, 2017b). Redistributive effects of nudges involve nudgees feeling morally licensed to engage in environmentally harmful behaviours or showing boomerang effects. A potential consequence of this issue is that GN do not deliver expected outcomes in diverse contexts (Hilton et al., 2018). For instance, studies often report a 'boomerang effect'; households that initially consume low energy or water tend to increase their consumption after being nudged through social norms (Allcott, 2011; Demarque et al., 2015; Schubert, 2017b). Boomerang effects are

[89] One reason for this success may be that countries with a marked concern regarding climate change demonstrate strong support for GN and other similar interventions (Sunstein et al., 2019).

observable when nudgees disprove a nudger's involvement in their choice contexts, thereby acting opposite to how the nudger expects them to.

Lack of trust in nudgers underlies redistributive response from nudgees. Schubert (2017b, 2014) notes that the perception of fairness of GN depends on how nudgees understand the intentions of the nudger in a democratic setting and whether they approve of it. Practices like enhancing transparency, accountability mechanisms, feedback, and engaging nudgees in the design of behavioural strategies, can build trust and subsequently minimise the instances of redistributive effects (and thus, unfair GN practices).

Another way in which GN can be unfair is by distracting nudgees from socioeconomic issues in institutions that contribute to climate change (Schubert, 2017b). This has implications for nudgee's prudence. In democratic set-ups, when nudgers, qua governments, exclude nudgees from the process of designing strategies for addressing climate change, a collective action problem, it takes away the opportunity for nudgees to learn about how their actions, opinions, and participation can influence environmental outcomes. For instance, GN using defaults are prone to such a charge. Ewert (2020) also observes, "...(to) be effective and acceptable, such approaches (to behaviour change) need to be built around co-production and a sense of partnership between state, individuals, and communities" (p. 555).

The risk of unfairness can be minimised, and prudence can be fostered in nudgees by making them more aware of the environmental impact of their actions (which is also a desirable systemic outcome). Moreover, nudgees can become more prudent by being able to assess the role and responsibility of nudgers in driving behavioural changes. This is also possible by introducing GN in a model of self-governance, a proposal considered in Chapter 7.

Focusing on trust provides the groundwork for understanding how GN systems can be better–one where nudges are effective and countervailing burdens experienced by the nudgee

can be minimised. Baier's (1986) view of trust being ubiquitous is useful for looking at how trust is fundamental for the smooth functioning of societies. However, since Baier (1986) focuses more on trust in moral interpersonal relationships as a way to protect vulnerability, her account does not extend to institutional systems, within which nudging operates.

Moreover, she acknowledges that her account of trust does not take into account trust between strangers and *networks* of trust (Baier, 1986). A network of trust encompasses multiple parties in a given context who share common expectations, overlapping domains of vulnerability, and responsibilities (Hardin, 2002, 2006). An interpersonal account of trust by Baier (1986), then, does not tell us much about how to develop trust in GN, which takes place in an institutional context.

Nudging involves networks, constituting a collective of nudgees and nudgers, and formal structures rather than individual trustees. To accommodate this, O'Neill's (2002f, 2018) account of trust in institutions is considered as a basis for an account of trust in GN. Such an account of trust, along with the two rules for building trust, is explored in Section 6.2.2. The next section outlines an account of institutional trust.

6.2. Trust and Trustworthiness: A Theoretical Framework:

In trusting nudgers, one does not trust a single individual, but a network and a hierarchical structure which is functional due to many individuals performing distinct roles in different capacities. In referring to GN, we are not concerned with placing trust in a person; we place trust in an institution, where different nudgers play roles that concern the design and deployment of GN. For instance, a social norm nudge being received through an energy provider will incorporate a data analyst who organises the information of thousands of households, a statistician or a researcher who interprets the data and designs a nudge, an administrative officer who approves the communication of a norm to a group of people. This

implies that nudgers can be individual actors who act as a part of an institution or a government, thus requiring an institutional account of trust, rather than interpersonal trust.

O'Neill (2002f) focuses on institutional trust, which is more complex than an account of interpersonal trust. It is easier to trust friends, family, or peers because we can easily find reasons to trust them (the evidence of their trustworthiness). The reasons to trust them could involve our past experiences with them, their commitment to us, their competence, or simply their honesty towards us. However, placing trust in others becomes complicated when we have to trust institutions that are formalised, or where our interaction has been remote or brief (O'Neill, 2002f, 2018). Thus, a theoretical framework for building and maintaining trust in institutions, as proposed by O'Neill (2002a, 2002f, 2018), is used to illustrate an account of trust in GN.

So far, two points are noteworthy. First, a morally justified form of trust between the nudgee and the nudger is required for GN to be justified from an ESC view. Second, as an account of institutional trust is necessary for nudging, O'Neill's account (2002a, 2018) offers a framework for evaluating institutional trust.

There is, however, a caveat: O'Neill is a deontologist. According to her account, within institutional contexts, placing trust based on the evidence of trustworthiness is necessary to uphold and protect the rights (or principled autonomy) of those who are vulnerable due to the actions of others (O'Neill, 2018, 2002e). O'Neill's (2018) normative account of trust is used in this thesis because, from both a deontological position and a system consequentialist point of view, trust is a functional resource necessary for the smooth functioning of societies.

Since trust has a functional value in leading to better systems, the sort of trust that upholds principled autonomy and fairness, and leads to better behavioural outcomes, is valuable. Hence, O'Neill's account of trust also holds value from a system consequentialist

standpoint by telling us how to place trust. This account, therefore, can guide us on how to build trust between the nudgee and the nudger within institutional contexts.

In what follows, the next two subsections explicate the nature of trustworthiness and a normative account of trust, outlining the two rules about trusting that will subsequently inform an account of trustworthiness and trust in GN.

6.2.1. Trustworthiness:

The question of when trust is justified inevitably leads to three concerns: who should we trust, in what ways, and why? According to O'Neill (2002a, 2018), trust necessitates trustworthiness. Placing trust in untrustworthy agents is risky; for it makes us vulnerable to deception or fear. Thus, it follows that we should only trust people who are trustworthy by assessing whether they are trustworthy or not to protect our entrusted matters.

Trustworthiness, when perceived by the trustor, demonstrates to them that it is safe for them to place trust in the trustee. Since trust is an attitude held by the trustor, certain properties, which are the properties of trustworthiness, justify that the trust is well-placed. For instance, in institutional contexts, one finds doctors trustworthy because our experience with them demonstrates that they are competent and committed to treating our health problems. Such aspects of the medical community lead to higher levels of public trust in doctors and physicians than in journalists and politicians, who are infamous for exaggerating claims and being involved in scandals. Simply put, trustworthiness involves the properties that can signal to trustors that a trustee is capable and willing to safeguard what has been entrusted.

There is a range of properties that are speculated to be signs of trustworthiness. For instance, researchers suggest that a symmetrical face appears to be more trustworthy than others (Todorov et al., 2008). Such traits cannot be a morally acceptable way of assessing

trustworthiness, for they focus on physical features, diverting attention from morally relevant properties in institutions like reliability and performance. We need a well-grounded approach to assess trustworthiness. The following subsection outlines different forms of institutional trustworthiness that are important in judging whether someone is trustworthy or not.

Truthfulness as Trustworthiness. As far as we expect someone to tell the truth, truthfulness is one of the ways to assess trustworthiness. Truthfulness is an *empirical* criterion for assessing trustworthiness (O'Neill, 2018). Within institutional contexts, being truthful can involve some extent of transparency, independent evaluations, and responsiveness. However, taking truthfulness as the sole criterion for determining trustworthiness is challenging in institutional contexts for two reasons.

First, placing trust based on truth claims is complex because evidence is often difficult to verify and open to falsification. In scientific enquiry, initial results and observations are rejected, reinterpreted, or refined so that scientific knowledge attains coherence, explanatory power, and simplicity (O'Neill, 2002f, 2018). This implies that an institution can be trustworthy even if it is challenging to verify truth claims.

Second, assessing the truthfulness of claims gets further complicated when we consider that distrust is often deliberately sown through conflicting claims. For instance, tobacco companies funded independent organisations to frame scientific evidence for smoking as an insignificant health risk (Drope & Chapman, 2001). Such examples show that it is difficult to verify truth claims in institutional contexts that are a part of politically and economically complex systems.

Thus, there must be other ways to assess trustworthiness, apart from truthfulness. O'Neill (2018) observes, "... to be trustworthy is to be trustworthy both in word and in deed, both in truth claims and in action." (p.295). Thus, O'Neill (2018) points out that speaking the

truth is not always enough for one to be trustworthy. For a trustee to protect the matters for which a trustor is vulnerable, one needs something apart from truthfulness as a feature of trustworthiness.

As a *normative criterion* for assessing trustworthiness, *Competence* and *Commitment* provide a feasible way to determine trustworthiness beyond the truthfulness of claims (O'Neill, 2018).

Competence as Trustworthiness. The trustee's competence in institutional settings is another way to assess trustworthiness. We trust others when we possess confidence that they are capable of protecting us in matters in which we are vulnerable. Competence can be interpreted in multiple ways: as an ability, as a skill set, as an expertise (Bennett, 2022; Mayer et al., 1995). However, for remote and distant interactions within institutional set-ups where we must place trust in people we do not know through personal interactions, we exhibit a domain-based form of trust (Holland & Stocks, 2017).

By exhibiting domain-based trust, we trust an institution, such as legal bodies, governments, and healthcare systems, for their area of competence. Since institutions involve a collective expertise, constituting a group of people where each is an expert in a field (medicine, education, finance), we trust with the expectation that a trustee, as a part of an institution, represents a collective that is competent regarding entrusted matters.

A domain-based account of trust more effectively captures how competence operates within collective and institutional contexts. There are two reasons for this. First, domain-based trust focuses on the competence of the trustee rather than their actions (Jones, 2019). We may continue to trust others due to their competence, even if they sometimes fail to deliver what is expected. Second, domain-based trust allows the trustee to make decisions that concern their expertise, even if our interactions with the trustee are remote. This implies that even if other forms of trustworthiness, such as truthfulness, are difficult to uphold, taking

competence as a sign of trustworthiness allows one to continue trusting despite the occasional setbacks (Holland & Stocks, 2017).

The primary consideration in expecting competence as a form of trustworthiness in the trustee is the credibility of the competence. Sporadic deception is prevalent, especially with the widespread use of digital communication (O'Neill, 2002a). Scammers and schemers like Harshad Mehta and Bernie Madoff use their skills and expertise to exploit or cheat people. O'Neill (2018) writes, "Those who trusted the well-known Mr Madoff with their savings could not console themselves for their losses…thinking that at least they had been trusting: far better if they had not been trusting…" (p.293). People can be experts in various domains, but some use their expertise to exploit people. We need to be wary of placing trust in them.

As truth claims are often difficult to verify and competence can be used at the expense of harm to a trustor (instead of benefiting them), another form of trustworthiness is needed to ensure that the trustee is trustworthy by being committed or willing to protect the matters of emtrustment.

Commitment as Trustworthiness. Commitment is a sign of trustworthiness, referring to the trustee's reliability regarding entrusted matters (O'Neill, 2018). By trusting the trustee's commitment, the trustor is confident that the trustee will consistently uphold the trust, despite changing circumstances. For a trustee to be trustworthy in one's commitment is to recognise and respond to the trustor's needs and expectations.

Commitment shows that the trustee is willing to, for whatever reason, uphold the trust that the trustor has placed in them. For instance, one might expect commitment from a physician because of professional standards or regulations that bind them. Similarly, a court shows commitment to the rule of law by upholding the rules of law even when there is pressure from political parties.

There are several ways in which commitment can be perceived as a form of trustworthiness. Within different theories of trust, commitment as a form of trustworthiness has been framed as responsiveness, accountability, or virtue of the trustee (McLeod, 2023; Potter, 2002; Holton, 1994).

Trust-responsive theory suggests that being trustworthy means that the trustee is responsive to the reason they must do something–all things being equal, they do it because the trustor is counting on them (Holton, 1994). Trustworthiness, as a commitment, may also mean that the trustee shares concerns or cares about the entrusted matters (McLeod, 2023). Moreover, one could also be committed to entrusted matters because of their "encapsulated interests" in the trust relationship (Hardin, 2006). Commitment could also result from one's virtue of being loyal to their profession, employer, or religious or non-religious ideologies (Potter, 2002).

For whatever reason one is committed, trustworthiness as a form of commitment requires the trustor to perceive the trustee as motivated or willing to take care of the entrusted matters. In being committed to the trustor's entrusted matters, the trustee puts forth the demand to deliver one's duties and secure institutional performance, irrespective of the reason why one is committed.

When viewed in the context of GN, these three types of trustworthiness, truthfulness, competence, and commitment, can provide ideas about how to build trust between the nudgee and the nudger, thereby justifying GN from an ESC perspective. The next subsection outlines the normative framework for building trust between two parties based on O'Neill's scholarship (2002a, 2002f, 2018).

6.2.2. Normative basis for trust:

While trust improves the outcomes of GN, it must be ethically justified because misplaced trust is dangerous. Trust placed by the nudgee must be reasonable; without trust, a

nudgee is prone to the risk of deception, loss of autonomy, and perceived unfairness. O'Neill's (2002a) normative framework of trust holds value from a system consequentialist standpoint because it tells us how to place trust carefully. If we appeal to the ubiquity of trust, then, much like any system that benefits from good trust practices, a system with GN would be better with the right kind of trust between the nudger and the nudgee.

Noting the need for rules about trust between the nudgee and the nudger, the two conditions from O'Neill's (2002a, 2002f, 2018) account of trust are explicated in the next two subsections. From her view, for trust between two parties to be morally justified, (1) the trustee must demonstrate trustworthiness, and (2) the trustor must possess the ability to assess trustworthiness.

Demonstration of Trustworthiness by the Trustee. Untrustworthy actors coexist alongside trustworthy actors in a particular domain, necessitating that institutions communicate their evidence of trustworthiness to the trustor. The trustee must demonstrate trustworthiness to the trustor, helping them identify trustworthy avenues. Signalling or showing trustworthiness can help the trustor to avoid deception by inviting trust in the right places (O'Neill, 2002a; Jones, 2012).

Information overload often makes it difficult for the public to perceive the truth. To demonstrate truthfulness[90], it is necessary to share knowledge and information in a manner that is "*accessible* to those for whom it might be valuable, *intelligible* to them if they find it, or *assessable* by them if they find and understand it." (O'Neill, 2018, pp. 298-299).

Similarly, demonstrating domain-related competence through good performance communicates the idea that the trustee can meet the trustor's expectations in a specific domain (Jones, 2012). To be trustworthy in one's commitment to the trustor is to recognise

[90] This point is revisited in Chapter 7. Accessibility and honesty in communicating GN are crucial for better integration of GN as legitimate tools for behaviour change.

and act according to the idea that the trustor is counting on one for specific reasons. Jones (2012) writes, “Trustworthiness is expressed in action when activated by being counted on” (p.72). To achieve this, a trustee can demonstrate commitment by being responsive, accountable, and forthcoming in matters entrusted to them.

Assessment of Trustworthiness by the Trustor. O'Neill (2018) writes, “Placing and refusing trust intelligently is demanding, and leaves opportunities for those who seek to push false claims, to fail in their commitments, or to pretend to competence and expertise that they lack.” (p. 295). Thus, the trustor must assess the trustworthiness of the trustee to minimise the chance of trusting in untrustworthy avenues. It is morally justified for a trustor to place trust in a particular party (or an institution) if they can assess whether the trustor is trustworthy in their truth claims, commitment, or competence.

One way to assess trustworthiness is to examine whether truthful claims are being made. Similarly, placing trust based on the evidence of credibility of competence allows the trustor to avoid the scope of deception and exploitation in trust relationships, especially in a time when recent technologies (digital platforms, social media) are positioned to bypass our capacities to judge standard competence through manipulation and exaggeration.

Although it is hard to judge commitment without past experiences with the trustee, O’Neill (2018) observes that, “...we need to focus not on generic attitudes but on ways to secure expert, professional and institutional performance without imposing ‘safeguards’ that distort that performance.” (p. 298). Judging commitment as a form of trustworthiness then involves looking for indicators of performance not based on attitudes, but through a history of performance, accountability measures, and responsiveness towards entrusted matters.

6.3. Trust in Green Nudges:

Recall from Section 6.1 that trust between the nudger and the nudgee is preferable for GN to be morally justified. While trust is important, misplaced trust can be dangerous. Thus, we must have a reasonable basis to trust nudgers qua institutions in charge of nudging.

As noted in Section 6.2, for trust between two parties to be justified in institutional contexts, the trustee must demonstrate trustworthiness. The trustor, in turn, must be able to assess it. Three forms of trustworthiness are acceptable in these contexts: *truthfulness*, *commitment*, and *competence*.

Given that trust is the basis for improving systemic outcomes and we have two rules for placing trust in a morally justified manner. The next question then is: What are the rules for placing trust in GN? To respond to the question, the section posits that two rules about trust are required: the nudger must demonstrate trustworthiness, and the nudgee must assess the nudger's trustworthiness. To support this argument, the next two subsections illustrate the normative criteria for placing trust in nudges. Moreover, potential ways in which trustworthiness can be demonstrated are also outlined.

6.3.1. Nudger demonstrates trustworthiness:

Why is a trustworthiness demonstration needed in GN? Since trust is an attitude (or a mental state), it is well-placed to the extent that the trustor perceives trustworthiness. To put it simply, trust is not only justified because the trustee is trustworthy and interested in the trustor's good, but it is well-placed when the trustor sees good reasons for placing trust. Thus, for trust to be reasonable between two parties, the trustee not only has to be trustworthy but also must come across as trustworthy to the trustor. For nudging, then, the institutions qua nudgers not only have to be trustworthy, but they also need to demonstrate their trustworthiness to the nudgee.

What happens when trustworthiness is not demonstrated?: In the absence of demonstration of trustworthiness by the nudger, it is possible that: (a) all the nudges, even the ones aiming to gain monetary ends, would come across as trustworthy to the nudge. Some agents leveraging behavioural insights may use brand value, advertisement, social media virality, or recognition. This can affect perceived trustworthiness (See Consumption Nudging (Braganza & Kapeller, 2024)), (b) the nudgees would not infer the value of eco-friendly actions, or (c) nudgees will not support the agenda of behaviour change at all.

In contrast to these three possibilities, from a system consequentialist perspective, it is desirable that a prosocial green nudge is identifiable and comes across as trustworthy for the nudgee who decides to cooperate with the nudgers by trusting them, and the nudgee recognises eco-friendly actions, thereby improving perceived legitimacy of nudges.

How to demonstrate trustworthiness?: While there can be intrinsic features of trustworthiness in a nudger, such as considering the mission of the institution as one's duty or being respectful towards the public opinion or feedback, the aspects of trustworthiness that the nudgee can perceive as signifiers of trustworthiness are interesting in this context. Thus, in what follows, the following subsections sketch how *Truthfulness*, *Competence*, and *Commitment* can be signalled to the nudgees.

Truthfulness and Green Nudging. There are several ways to argue that truthfulness is a criterion for judging a trustee's trustworthiness in institutional contexts. Truthfulness can be seen as a sign of personal or professional integrity and as possessing character. One of the obvious ways for a nudger to be truthful in institutional contexts is by sharing the right information openly. The upside of making truthfulness a requirement is that it reduces the tendencies and the inclination to deceive or to cheat (O'Neill, 2002f). O'Neill (2002a, 2002f) argues that the fear of being scrutinised or held accountable due to mandatory openness keeps public authorities and institutions in check. The downside of truthfulness is that the

perception of honesty or openness depends on how the trustor perceives it. Subsequently, for the nudgee to perceive the information as truthful, truthfulness can be demonstrated in nudging by maintaining transparency.

Transparency. The debate on transparency[91] in nudges has gathered traction in the literature. The level of transparency refers to the extent to which information can be openly shared with the public. In the context of nudges, a satisfactory level of transparency requires that the nudger make a reasonable effort to inform the nudgee that their behaviour is being influenced (Santos Silva, 2025).

Transparency in decision-making processes and the outcomes pursued are perceived as indicators of trust within institutional and policy contexts (Brezzi et al., 2021). According to Thaler and Sunstein (2008), the requirement for transparency satisfies Rawls' Publicity Principle, which suggests that policies should be defensible in public.

In considering the extent of transparency, it is suggested that a nudge is sufficiently transparent if the target groups of nudgees can identify the goals (type interference transparency) and the mechanisms of nudging (token interference transparency). Otherwise, it is a non-transparent nudge (Hansen & Jespersen, 2013; Ivankovic & Engelen, 2019; Bovens, 2009).

A more nuanced view suggests that nudges must be transparent to the extent that their goal and mechanisms can be detected by a 'watchful' nudgee who can blow the whistle on intrusive nudges (Bovens, 2009; Ivankovic & Engelen, 2019). It is important to consider the

[91] Transparency has implications for the nudgee's autonomy. Santos Silva (2025) suggests that transparency is important because it triggers nudgee's engagement with System 2, the reflective, deliberate, and analytic cognitive processes. As this allows nudgee the opportunity to engage more critically with nudges, transparency preserves their autonomy (Santos Silva, 2025). This also means that nudgee can engage with the proposal of behaviour change, be critical about the methods of nudging, all of which is desirable for upholding principled autonomy.

extent of ‘watchfulness,’ which corresponds to alertness or awareness of a nudgee, because increasing transparency of the nudges fails its purpose if a nudgee cannot detect a nudge.

There are three reasons why transparency is necessary for establishing trust. First, openness in practices that seek desired collective outcomes promotes trust. It makes intuitive sense that transparency increases the perceived legitimacy and acceptance of an intervention that promotes a desired goal. For instance, in a country with a strong consensus about climate action, implementing GN is seen as a welcome move (Sunstein et al., 2019). Moreover, in recent years, the focus on communicating information openly or transparently to the public, interest groups, and campaigning organisations for social goals has become a part of the openness agenda (O’Neill, 2002f).

Second, transparency is crucial for truthfulness and trust, as it demonstrates the congruence between the trustor’s values and the trustee’s values, consistency, and fairness (Mayer et al., 1995). If a nudgee trusts the nudger, they will not only follow the direction of a nudge but also see how nudges support their goals (Ivankovic & Engelen, 2019).

Third, transparency is needed for nudgees to identify the nudges that pursue collective goals. Given that both private and public actors use nudges to influence behaviour, it is reasonable to assume that transparency can help nudgees identify illegitimate nudges (those seeking monetary ends) (Lenzi, 2019). Transparency is required to distinguish GN from illegitimate practices, such as ‘green washing’ and ‘green dark patterns’, which seek to mislead individuals[92]. Santos Silva (2025) outlines the two concepts in this context: The European Union Agency for Fundamental Rights defines ‘green washing’ as misleading and false environmental claims about products having no harmful environmental impacts. 'Dark green patterns’ refer to the deceptive behavioural design practices on digital interfaces that

[92] Another such concept is ‘Consumptive nudging’, which can be seen as an umbrella term for all these practices (Braganza & Kapeller, 2025).

make it difficult for one to avoid behaviour change by imposing additional fees, false claims, or confusing UX (User Experience) design. As GN operate in the same context as these deceptive practices, nudgers have a reason to make GN transparent so that nudgees can identify and avoid greenwashing and dark patterns (Santos Silva, 2025).

The goal of making GN transparent is so that a nudgee perceives the presence of a nudge and understands the pro-environmental behaviour promoted by it. It is better if a transparency GN creates awareness and encourages nudgee to engage with System 2 to understand their behaviour. Due to these three reasons, from a system consequentialist perspective, a nudge that is transparent and considered truthful by the nudgee is better than a nudge that is non-transparent or has an insufficient level of transparency.

An objection is that nudges may work adequately because of their covertness and, therefore, a lack of transparency (Bovens, 2009). It is, however, contested in evidence whether increasing transparency compromises the efficiency of nudges (Santos Silva, 2025). The extent to which transparency compromises the performance of a GN is open to further research and discussion, which may be an issue with this feature. Noting this limitation, two other bases for assessing the trustworthiness of nudges are considered, namely, competence and commitment.

Competence and Green Nudging. Nudgers should be competent because their competence influences our trust in them. From a philosophical perspective, in being competent, one seeks to be free of errors in delivering what is expected of them. From a psychological standpoint, evidence of competence allows our trust in others to flourish and increases the extent of cooperation (Mayer et al., 1995). Moreover, institutional trust results in a higher level of cooperation from the public (Brown, 2020), a property that can further improve the effectiveness of GN.

Demonstrating evidence of competence is critical for building an attitude of trust, as the trustor requires evidence of a trustee's competence to evaluate their trustworthiness. Selinger and Whyte (2011) argue that the "moral acceptability of nudges hinges in part on whether they can provide an account of the competence required to offer nudges, an account that would serve to warrant our general trust in choice architects." (p.461). Considering this, the following signifiers of competence and trustworthiness in nudgers can be outlined.

Performance. For a nudger to be competent, designing and implementing nudges effectively and efficiently is needed. To evaluate the performance of public policy systems properly, efficiency and effectiveness are two essential factors (Mergoni & De Witte, 2022). Efficiency assesses a system's ability to achieve its stated outcomes. Effectiveness refers to the output of the system in terms of the intended outcomes (Mergoni & De Witte, 2022). These two categories for performance evaluation can be extended for assessing the competence of nudgers.

In the context of GN, efficiency typically refers to how GN perform in terms of the traditional cost-benefit analysis (Santos Silva, 2025). It involves analysing quantitative data to assess and predict whether a GN strategy can reduce resource and time-related costs associated with behaviour change. The efficiency of GN is not a sufficient measure to ensure performance. For instance, an efficient green nudge may save resource costs in the short term, but it may not lead to long-term behavioural shifts. Thus, considering qualitative aspects related to performance, i.e., effectiveness, is an important basis for nudgers to design GN competently.

Santos Silva (2025) notes that effectiveness, in the context of GN, is a spectrum. To say that effectiveness is a spectrum means that " an intervention will be effective if it is better than the previous state of affairs" (Santos Silva, 2025). It questions whether GN leads to pro-environmental behavioural outcomes, while considering the qualitative costs (Santos Silva,

2025). This means that the effectiveness of GN is concerned with whether they can bring about a behaviour change, while also justifiably reducing countervailing costs, such as the costs of autonomy, privacy, and spillover effects (the boomerang effect (6.1.3) and moral licensing (4.1.8)), in the long term.

In other words, from a cost-benefit analysis perspective, GN must lead to pro-environmental outcomes without posing undue social, cultural, psychological, and ethical burdens on the nudgee. Undue burdens are posed in situations where the benefits of GN on third parties outweigh the costs (such as inducing guilt, social pressure, or loss of autonomy) on the nudgee (DesRoches et al., 2023).

Additionally, moral intent is not necessary for GN to be effective. This suggests that GN can be considered adequately effective even when they improve behavioural outcomes due to unintended or unrelated reasons. Santos Silva (2025) provides an example of an energy company that nudges consumers to reduce energy consumption during peak hours. Although the nudge can offer financial incentives to reduce energy usage, it still qualifies as a green nudge due to its pro-environmental and socially beneficial effects.

Privacy Measures. The ease of technological advancements in GN comes at the cost of users' informational privacy[93] (data related to race, age, income, gender), patterns of behaviour, and preferences being accessible to those who may intend to exploit this information for various purposes (monetary or political interests). Such data are used in

[93] The information about one's work, health, and relationships has become critical for self-management, social interaction, and communication through technology. Datafication involves using such information gathered through smartphones and wearable devices for monitoring, processing, performance, and often monetization. With the assistance of datafication through new technologies, nudges are deployed more on digital platforms, which are privacy-sensitive contexts. This is to say that user data acquired on such platforms is often threatened by the interests of third parties or big-data barons (data companies like Meta and Google) for monetary ends (Yeung, 2019).

nudges to create personalised defaults, tailor social norms, or appeal to the nudgee's self-image, thereby challenging informational privacy as a result.

A common example of this is nudges, which fix a default choice by learning from users' previous preferences, or smart meters storing and analysing user history to send social norms (Hilton et al., 2018; Sunstein & Reisch, 2014). A trade-off needs to be made: to nudge better, we need more data about the nudgee, but access to more user data for different nudgers raises concerns about privacy protection. A threat to privacy compromises the principled autonomy of the user (qua nudgee) and challenges the ethical permissibility of GN. There may be an ethical trade-off between effectiveness and informational privacy. However, we must consider effectiveness in terms of long-term behavioural outcomes, improved environmental preferences, and reduced ethical risks to the nudgee. The ethical justification of GN, in this case, requires us to look beyond the short-term gains possible by risking privacy.

There are three approaches to addressing privacy concerns in nudges. One way to safeguard user privacy is to improve informed consent practices, establish robust information regulatory mechanisms, and carefully set up default nudges. A second approach is to design GN for preventing environmentally mediated harms, as well as improving privacy practices, something like 'green-privacy nudges'. The use of nudges to raise awareness about user privacy is a much-discussed topic in the literature. For instance, it is possible to learn from visual nudges, which have shown promise in encouraging people to avoid privacy-invasive practices (Choe et al., 2013).

A third approach is to establish robust legal mechanisms to ensure that there are 'legitimate interests' in storing the personal data of the nudgees, particularly in the private sector (Santos Silva, 2025). For instance, GN can be used to collect and process information, such as the number of people in a household, energy/water consumption, age, and

demographic factors. However, if these data are either published or sold to third parties, the use of GN would no longer be considered legitimate (Santos Silva, 2025; Ranchordas, 2019). Such instances would involve violations of data protection laws (or GDPR in the context of the European Union).

In summary, one of the measures to design GN competently involves GN upholding informational privacy, which may involve ensuring legitimacy and raising awareness about privacy-sensitive contexts.

Adaptability. Adaptability is a signifier of competence. Being adaptable means responding to changes by adjusting one's approach to accomplish the required outcome. This skill is especially needed when GN operate within heterogeneous or diverse environments.

Adaptiveness involves understanding individual reactions to the GN and the direct and indirect effects of a nudge on a particular group. Understanding such factors is important for securing public support in favour of behaviour change. Various studies corroborate this insight. A recent study demonstrates that the effect of social norm nudge on the pro-environmental decisions of some nudgees depends on their existing preferences (Picard & Banerjee, 2025). For instance, the effects of the nudge varied among individuals who were transitioning to vegetarianism, those who were already vegetarian, and those who occasionally preferred a vegetarian diet (Picard & Banerjee, 2025).

In nudging political liberals and conservatives in the USA, the framing of nudges affects whether a particular group responds positively or negatively to them. Tannenbaum et al. (2017) found that designing nudges around liberal or conservative policy goals has a bearing on the success of nudges. Nudges are considered more ethical by nudgee qua respondents when they align with their political views. In the study, the authors recommend that the nudges pursuing collectively desirable outcomes should not be implemented,

considering their face value, because nudgees tend to conflate their attitudes about politics with their attitudes about policy methods[94] (Tannenbaum et al., 2017).

To design a more adaptable GN, an efficient strategy would be to tailor it based on the user profiles, as the impact of a nudge varies depending on the group being nudged. To avoid compromising the effectiveness of nudges in the heterogeneous groups, Picard and Banerjee (2025) recommend designing GN in a way that does not crowd out further engagement with other nudges. Additionally, to achieve adaptability, nudges must be implemented considering contextual factors, such as semantic variance[95] and behavioural biases[96] within subgroups.

[94]A counterargument in this context may question the possibility of a cognitive shift in such instances, a systemic goal outlined in Chapter 5. A cognitive shift, constituting the shaping of social norms and the internalization of moral obligations, may appeal differently across social and political groups. Since nudges and behaviour interventions require voluntary participation, fundamental political differences in how individuals perceive nudges are often unavoidable. This is not to say that GN cannot lead to cognitive shifts. Rather the method of nudging, and therefore the core values it appeals to in each group, can be different. So, for a conservative, the reason to follow their moral obligation through GN can be different from that of a liberal. Nevertheless, in each of these cases, GN can support the objective policy goal that maximizes positive systemic outcomes.

[95] To use nudges successfully, the nudger ought to understand the meanings that are embedded in the daily experiences of the nudgee. They must attend to the diverse perceptions of meaning, tied to differences in identity and context, that influence how users interpret nudges, what is also called semantic variance (Selinger & Whyte, 2010, p 436).

[96] Similarly, biases are observed differently across cultures (Bovens, 2009). He proposes, "Biases may interact with culturally variable features and hence will meet with different cultural implementations (in nudges)" (Bovens, 2010, p.483). Two types of biases influenced by the cultural context are *Cognitive* and *Behavioural* biases (Bovens, 2010). *Cognitive* biases refer to the idea that certain norms and attitudes are embedded in one's cultural experiences. Similarly, *behavioural biases* refer to cognitive traits such as the endowment effect or loss aversion, which are shown differently across cultures. The presence of these biases and their varied effects on different population groups suggests that nudgers must take into account these biases before designing nudging strategies.

Explainability. Another way to demonstrate competence and invite trust from nudgees is by explainability. Santos Silva (2025) writes, "A nudge is explainable if the policy objective it pursues and the means by which it operates are known, or reasonably ought to be known, to the person being nudged." (p.20). Transparency complements explainability because it determines the extent to which the nudge is communicated openly to the nudgee.

Explainability in GN requires going beyond making nudges transparent. It involves framing nudges in a way that makes it more likely for the nudgee to perceive the intent, goal, and process of behaviour change. So an explainable nudge is more than transparent. It appeals to the reasoning capabilities of the nudgee, invites them to deliberate, and attempts to preserve the effects of a nudge for the long term[97].

Two types of nudges are notable in this context: Nudge plus and Boosts. Nudge plus incorporates a reflective strategy into the design of nudges that persuades a nudgee to think about their decisions (Banerjee & John, 2021). Boosts target the competencies of a person by improving their decision-making abilities, knowledge, and raising moral concerns (Hertwig & Grüne-Yanoff, 2017). Moreover, a combination of nudging with boosting is suggested as a promising approach to improve citizen capabilities (DellaValle & Sareen, 2020). In short, both nudge plus and boosts are explainable nudging strategies, as they try to instigate a more meaningful behaviour shift.

Commitment and Green Nudging. Philosophically, demonstrating commitment by acting in a particular manner signals a virtue expected in the institutions that act for the public good. If the trustee demonstrates a commitment to the reasons the trustor is vulnerable or the goals they desire, it improves the trustor's likelihood of trusting them. Commitment in nudging warrants the nudger responding to the nudgee's concerns regarding their

[97] These insights are developed in the chapter on recommendations (Chapter 7).

environmental behaviour and seeking the long-term accomplishment of environmental goals (as opposed to short-term benefits). Considering this, some ways in which the nudger can demonstrate commitment as a form of trustworthiness to the nudgee are outlined.

Responsiveness. One way to demonstrate commitment is by being responsive to what the nudgee expects the nudger to do. An institution's responsiveness positively correlates with trust (Kavanagh et al., 2020). The responsiveness of an institution can be defined as the ability to provide accessible, efficient, and citizen-centred public services that effectively address the public's expectations, preferences, and requirements (Brezzi et al., 2021). In GN, the responsiveness of the nudger can demonstrate that the nudger is trustworthy. At least two steps can signify responsiveness in nudging contexts. First, removing or modifying ineffective GN practices based on monitoring and public feedback. Second, responding to public concerns regarding the use of nudges through public assemblies, citizen engagement platforms, and participation forums.

Accountability. Another way to demonstrate commitment to the nudgee is by upholding accountability, a feature that is often found to be lacking in the nudgers. A lack of accountability reflects on the nudgee as if the nudger is ignorant of incorporating citizenry in designing and strategising behavioural shifts. For instance, a lack of accountability is demonstrated when a nudge is considered to have a suboptimal level of transparency by the nudgee (Ivankovic & Engelen, 2019).

Being accountable entails a moral obligation to deliver what is expected of us (Talbert, 2019). Accountability structures are necessary for building trust in nudging because they can transfer the responsibility of maintaining sufficient transparency, monitoring, improvement, and dealing with uncertainties. This means that as nudgers, institutions need to be accountable by maintaining a specific code of conduct on all levels, handling uncertain situations and spill-over effects, ensuring competence (through performance metrics or audits), and maintaining fairness (Talbert, 2019).

In designing and implementing GN interventions, nudgers may demonstrate accountability in at least three ways. One way is to maintain sufficient levels of transparency,

as already indicated. Another way to build accountability is to use participatory approaches by combining nudging with public deliberation, which can pave the way for long-term behavioural changes to address climate change, a proposal also considered in Chapter 7 (Lenzi, 2019; Talbert, 2019). Thus, building accountability this way not only demonstrates trustworthiness but also leads to long-term benefits from GN, which is desirable from a system consequentialist perspective. Lastly, structuring institutional mechanisms to address oversight and ensuring fairness in nudging can build the perception of accountable institutions.

Value Commitment. One way to demonstrate commitment to a cause and be perceived as trustworthy for the same reason is by showing that one's values align with those of the trustor. As humans operate with limited epistemic knowledge about the world, it is not difficult to see how personal preferences, values, beliefs, and motivations play a critical role in determining whether one can trust another (Baier, 1986).

The extent to which an individual can trust another depends upon the trustor's perceived trustworthiness of the trustee. Thus, when assessing the trustworthiness of the trustee, a trustor also considers whether the trustee's recommendations align with what is important to them and accord with their values (Bennett, 2022). This is also an important dimension for adaptability, and thus for communicating the trustworthiness of the institution. This means that nudges must be designed to pursue systemic goals, but at the same time, their framing can be such that a nudgee is convinced and supports the behaviour change because it aligns with their value systems.

Framing nudges to appeal to the values and preferences of diverse groups can lead to long-term performance improvements. This may seem difficult to do, considering the diverse contexts in which people hold uncertain preferences and different ideologies. However, this is not the case. Through a study on voting patterns in conservative and liberal groups, Feldman

(1988) observes that achieving social goals is possible by structuring messages to tap into core belief systems and values. The challenge posed by climate change to economic individualism, family systems, and free enterprise would appeal to conservatives. Similarly, appealing to human rights, issues of equity, intergenerational responsibility, and global cooperation would tap into the core values of liberalism (Feldman, 1988). A similar insight is reflected in Tannenbaum et al. (2017)'s work on framing nudges differently for conservatives and liberals.

Thus, learning from these studies, it can be inferred that in the absence of common ideological preferences, nudges can be tailored to appeal to specific targeted groups. While these are some of the ways in which the nudger can demonstrate trustworthiness, the next step in building trust between the nudgee and the nudger is to find ways in which the nudgee can perceive the evidence of trustworthiness.

6.3.2. Nudgee assesses trustworthiness:

Why is a trustworthiness assessment needed? Using subconscious processes to influence behaviour implies that a nudgee often does not perceive the nudges actively. Recall that trust is an attitude. Trust can be perceived by nudgee only when it is demonstrated enough so that the nudgee can adopt an attitude of trust towards the nudger. Apart from demonstrating trustworthiness, it is possible to increase the chances of nudgees to perceive trustworthiness by building their capacities to grasp and assess the evidence of trustworthiness.

Thus, along with the demonstration of trustworthiness, perceiving the evidence of trustworthiness by the nudgee is another condition. Note that while assessing trustworthiness can build trust, the moral justification requires the nudger to possess the aspects of

trustworthiness perceived by the nudger that warrant trust[98]; that is, the nudgers do not pretend to have the trustworthiness that they do not possess.

What happens when the nudgee does not assess trustworthiness? If a nudgee does not assess trustworthiness, then it is possible that (a) they do not actively try to make sense of their behaviour under the influence of nudges and do not understand and internalise their moral obligations, (b) they either continue to rely on nudges to make decisions, or (c) oppose the use of nudges. There is another possibility that (d) the nudgee misplaces their trust in a trustor who intends to leverage their behaviour for monetary gains.

Recall that an activity concerned with a system is morally justified when it is according to the rules that lead to the best possible outcomes for that system. In comparison to the possibilities (a) to (d) discussed above, GN can lead to a better state of if they result from rules that lead to a nudgee (a') trying to make sense of their behaviour and developing eco-friendly habits, (b') continues to act in eco-friendly ways even in absence of a nudge, (c') supports the use of behavioural insights or nudges to promote environmentally benign outcomes, and (d') places trust carefully in the nudgers that genuinely seek pro-environmental ends. The situations (a') to (d') are possible when nudgee assesses the evidence of trustworthiness of the nudger. Thus, from a system consequentialist point of view, a nudgee should assess the evidence of trustworthiness of the nudger.

How to assess trustworthiness? One of the ways in which nudgee can assess trustworthiness is by being 'watchful.' Recall that a nudge was argued to have a level of transparency that is detectable by a watchful nudgee (Ivankovic & Engelen, 2019). Although authors have developed the notion of watchfulness for assessing the transparency of a nudge, there is a possibility that watchfulness can help one determine whether a nudge is designed

[98] Pretending to have an expertise that one does not have is a case of deception, as we invite others to be vulnerable to our actions without actually possessing the skillset or the ability to protect them regarding those matters.

with competence and demonstrates commitment. Normatively, four conditions are needed for a nudgee to be watchful: (a) educational, (b) legal, (c) democratic, and (d) societal (Ivankovic & Engelen, 2019).

The first condition is that a nudgee needs to be educated about how behavioural insights work and to what extent. Although not an extensive requirement, it demands that a nudgee understands the subconscious behaviours and how they influence their daily activities. There are three ways to do so. First, combining nudges with other strategies to help nudgee reflect on their behaviour, for instance, nudge plus and boosts. Second, prompting nudgees to incorporate behavioural changes by working with their habits. This may involve analysing their daily behaviour and tailoring nudges based on their routine. Third, highlighting the need for behaviour change in climate campaigns, appealing to individual obligations to address climate change).

The second condition requires legitimising nudging through rules, mandates, and legal mechanisms. This argument is made by Lepenies and Malecka (2015), who argue that incorporating nudging and behaviour change into legal systems can serve as a safeguard to minimize the chances of behavioural insights being misused.

The third and related condition requires that a nudger declare publicly their intentions, aims, and the mechanisms for using behavioural insights to change behaviour, a requirement already stipulated by Thaler and Sunstein (2008) through Rawls' Publicity Principle.

The fourth condition requires certain nudgees with expertise in particular areas, such as health, climate change, and finance, to assess the nudging strategies and provide feedback through democratic processes (Ivankovic & Engelen, 2019). This is possible by creating democratic spaces around behaviour change narratives, involving citizens in the behaviour change process, and possibly by introducing polycentricity in behavioural governance. In

what follows, the next chapter conceptualises Behavioural Self-governance (BSG) that can play a role in achieving these conditions.

6.4. Conclusion

This chapter outlines an account of morally justifiable trust between the nudgee and the nudger. There can be certain objections to this account. First, institutional trust is complex, and the trustworthiness features mentioned in the chapter may not be sufficient for a policy-level implementation of GN. Policy scenarios are complex, where preferences vary across groups, and thus, different groups of nudgees can have varying degrees of trustworthiness. However, even though real-world trusting practices and mechanisms may vary, the rules about trust and trustworthiness in this chapter can provide a template.

Two rules, 1) demonstrating trustworthiness by the nudger and 2) assessing the evidence of trustworthiness by the nudgee, provide an important starting point for building trust between the nudgee and nudgers. Moreover, a better understanding of trustworthiness features, such as improving performance, adaptability, explainability, and responsiveness, can pave the way for more effective empirical research on the topic, resulting in trustworthy nudging practices to address climate change.

Second, the focus on institutional trust may be challenged by pointing to other ethical concerns related to nudging. For instance, the second objection can be framed as: While trust between the nudgee and the nudger is promising for the GN to lead to successful outcomes, is the instrumental value of trust sufficient for moral justifications of GN?

There may be other values that are necessary for the moral justification of GN from a system consequentialist perspective (or ESC, considering ecosystemic outcomes). However, trust is a pivotal, functional, and non-negotiable value for justification from this view. As discussed in Chapter 5, certain objective values are intrinsically valuable for good systemic

outcomes, such as prudence, autonomy, and fairness. There can be other objective values, such as dignity and civic virtues like solidarity, or personal values like temperance, that are crucial for GN.

So far, the thesis has focused on building a moral defence of GN by arguing that GN is a tool that helps us fulfil our environmental obligations, is justifiable from a systemic view, and more so when it is functional through institutional trust. The next chapter shifts the focus from the moral justification of nudges to the challenges policymakers face in practice. To address these challenges, Chapter 7 proposes a model of self-governance for GN, referred to as Behavioural Self-Governance (BSG), amongst other recommendations.

Chapter 7. Behavioural Self-Governance and Other Recommendations for Green Nudging

As outlined in the previous chapter, trust is a functional value for upholding other objective values. This requires effective communication of trustworthiness from the nudger and the ability to assess the evidence of trustworthiness from the nudgee's end. In practice, building institutional trust[99] depends on public acceptance, engagement, and cooperation with the proposal for behaviour change, a point also noted by Grelle et al. (2024) and Schubert (2017b).

The discussions in this chapter address policymakers and readers who agree that trust between the nudgee and the nudger is a practical proposal, along with the signifiers of trustworthiness, yet nonetheless ask: How can we create the conditions for morally justifiable trust, and therefore ethical GN, to emerge in practice? To this end, this chapter offers three procedural recommendations for integrating GN in human-centric systems.[100] While this is a broad and ambitious proposal, three procedural strategies are developed herein: Honest framing of behavioural goals (7.2), Regulation of Consumptive Nudging (7.3), and Behavioural Self-governance (7.4).

7.1. The Need for Procedural Recommendations:

The critique of nudging (devices) comes from three broad academic camps: ethical, societal, and cultural (Hukkinen, 2016). So far, this thesis has focused on ethical critiques, the effectiveness, ethical defence, and the conditions under which GN is ethical. Societal and cultural considerations are integral to GN's ethical legitimacy because the context in which

[99] The ethical argument, so far, has been the following: We have individual moral obligations. GN helps in fulfilling these obligations. Even though they are effective, GN raise specific ethical issues. GN can be justified according to a system consequentialist position if they operate according to rules that maximise good ecosystemic outcomes while upholding specific objective values. Trust between the nudge and the nudger is a functional value for upholding these objective values. Chapter 6 sketched an account of a morally justified form of trust between the nudgee and the nudger.

[100] Doing so echoes Singer's (1993) point about the nature of any ethical enquiry, "…: an ethical judgement that is no good in practice must suffer from a theoretical defect as well, for the whole point of ethical judgements is to guide practice" (p.03).

GN operates is deeply interconnected with these systems. Singer (1993) also notes, "The consequences of an action vary according to the circumstances in which it is performed." (p.03). This means that these systems are an important consideration, as they shape immediate and long-term consequences associated with GN. Noting this, the following chapter engages with the circumstantial considerations that are crucial for GN to be ethically justifiable.

Each recommendation in this chapter concerns the social and/or cultural critique of nudging. The main social critique of the state-of-the-art nudging is that it is undemocratic and elitist, and that it ignores the need to empower citizens to participate in decision-making processes that affect their welfare (Hukkinen, 2016). These concerns are also reflected in the mainstream political critique of nudging (Viale, 2016; Button, 2018)

For example, Selinger and Whyte (2011) argue that focusing on behavioural changes without legitimate mechanisms is a slippery slope because it bypasses citizens' reasoning abilities, which are crucial to democratic societies. Similarly, Button (2018) and Bovens (2009) argue that the government's approach to designing and implementing behavioural strategies, such as nudges, may create an unskilled citizenry due to a lack of informed preferences. Button observes, "..this (nudging) is a targeted form of secular redemption that does not depend on the individual's soul, mind, or even self-awareness: nudges constitute 'reason by-passing non-argumentative influences.'" (Button, 2018, p.1037).

Schmidt (2017) also argues that the power to change behaviours results in relations of alien control[101] — a position in which some people can impose their will and decisions on others. This provides an entry point for those seeking monetary ends (Schmidt, 2017). Most scholars argue against considering nudgees as passive actors in the behavioural change

[101] Schmidt (2017) defines alien control as, "An agent A holds alien control over someone else B, when A has the power to influence B's life and such power is not controlled by B (B is an individual, but A can include individual and collective agents)." (p.404).

discourse. The main idea is that such an approach depoliticises citizens and exposes them to corporate preferences, and reduces their democratic engagement with climate change (Button, 2018; Ewert, 2020). Moreover, this critique also resonates with Chater and Lowenstein's (2023) concerns regarding nudging being limited to i-frames. The depoliticisation of citizens becomes a risk when the focus is only on targeting individual behaviour and not the broader political and social context that requires behaviour change.

The cultural critique of nudging considers its cultural and contextual insensitivity (Malecka, 2021; Hukkinen, 2016). Policymakers often find it difficult to tailor strategies to dynamic contexts due to their limited understanding of the nuances and intricacies embedded in socio-cultural contexts (Varazzani & Hubble, 2024; Malecka, 2021).

For instance, to be culturally relevant, nudgers need to consider semantic variance, an effect in which diverse groups of people perceive meanings differently (Selinger & Whyte, 2010). A similar phenomenon is cultural variance, suggesting that the norms and messages communicated by the policymaker through nudges must take cultural intricacies into account to adapt nudges effectively to a given context[102] (Santos Silva, 2022; Bovens, 2010). The cultural contexts are especially important, given the potential for integrating nudging with algorithmic processes, which can exacerbate existing concerns about algorithmic bias[103] (Mills & Sætra, 2022; Sunstein, 2023).

As both societal and cultural critiques of nudging impact how effective GN is and its legitimacy in democratic societies, an examination of GN, then, to some extent, needs to engage with these considerations. In what follows, three recommendations that address the

[102] An example (although not specifically of green nudging) is the failure of nudges implemented by the Obama Administration to promote handwashing and hygiene to address a health crisis in Flint. The behavioural strategy resulted in public backlash and administrative failure. The strategy failed because it ignored the town's "racial history, its socioeconomic circumstances, all of it" (Małecka, 2023, p 351).

[103] The set of ethical issues for the artificial nudgers (using GN) is a proposal that warrants a distinct but related examination. Such an ethical examination will have to consider the risks and scope of Artificial General Intelligence (AGI), explainability and trust in algorithmic systems.

social and cultural critiques are taken up. Broadly, these recommendations rest on conditions that are likely to be foundational for any successful policy.

Two recommendations emphasise that nudgees should be seen as active recipients of pro-social ends, and they should be engaged with the goals and processes of achieving climate outcomes, including honest framing of behavioural goals and Behavioural Self-governance (BSG) (7.2 and 7.4). To be an active recipient means that nudgees should be seen as autonomous agents who are capable of being engaged in the process of designing and shaping choice architectures. These two recommendations are concerned with improving the ability of nudgees to assess the evidence of trustworthiness and the expectation from the nudger to generate such conditions. They appeal to the idea that by including nudgees in the design process and treating them as democratic agents, GN can be better positioned to address climate change (Banerjee & John, 2021; Harmon-Jones et al., 2013; Chater & Loewenstein, 2021).

Another recommendation is directed at nudgers and governments who are in a position to do to limit Consumptive Nudging (7.3) (Braganza & Kapeller, 2025). The goal of this recommendation is to increase the prospects of positive behaviour change for the nudgees. This can be done by limiting the power of the private companies to nudge in uncontrolled, opaque ways (Schmidt, 2017)

7.2. Honesty and Accessibility of Green Nudges:

Honesty and accessibility regarding the goals of GN are crucial[104]. The extent of accessibility[105] in the framing of GN goals has implications for its effectiveness. Grelle et al. (2024) note that the public's reception of GN depends on how policymakers frame the goal of GN (as personal preference or as a societal good). Schubert (2017b) writes that the ethical

[104] In the previous chapter, truthfulness as a form of trustworthiness was discussed. The discussion here is different because it is concerned with the implications of honesty for effectiveness, along with trust.
[105] Accessibility is concerned with how understandable, transparent, and relatable GN is for the nudgees.

quality (or assessment) of a GN strategy "will always depend on intricate value trade-off whose solution, in a democratic setting, is ultimately up to the citizens" (p.336). This means the nudgers should carefully consider how they present GN's goals, as this affects how acceptable and effective the GN is to the nudgee.

To illustrate this point, consider the example of indirect water conservation studied by Goldstein et al. (2008). A common type of GN involves communicating descriptive norms to hotel guests, encouraging them to reuse their towels rather than have them changed daily. To avoid unnecessary laundering, hotel staff often leave out notes for the new guests, indicating the overall percentage of guests who reuse towels. This practice confers environmental value on a particular action, encouraging guests to act in ways that communicate pro-environmental behaviour to themselves and others.

In this particular case, hotel guests changed their behaviour because the descriptive norm tapped into their self-image as pro-environmental actors and/or their social image as eco-conscious citizens (signalling their behaviour to others). Goldstein et al. (2008) observed that communicating an inflated figure about guests reusing the hotel towels to the nudgee led to a significant behavioural change[106].

If the information communicated by the nudger is identified as misleading, this signals to the nudgee that the nudger may be willing to lie[107]. This undermines the nudger's credibility and sets a precedent that affects trust in GN.

Even if a nudgee accepts the information as truthful and acts according to it, this still has implications for the integration of GN as a strategy to tackle climate change. Although GN immediately improve the behavioural outcomes in such cases, internalising false

[106] Schubert (2017b) notes such practices as an 'obvious' ethical concern. There are two reasons why being dishonest about the shared information is an ethical concern. First, it compromises prudence — the ability of the nudgee to foresee the environmental impact of their actions, deliberate about it, and internalise the right environmental actions. Second, as outlined in Chapter 6, nudgees find nudgers to be trustworthy if they assess their claims as truthful.

[107] This also has implications for the perceived democratic acceptability of GN.

information and norms can undermine efforts to cultivate environmentally conscious citizens. Despite short-term effects, over longer periods of time, people may struggle to improve their environmental actions if repeatedly exposed to false information. All things being equal, addressing climate change and fulfilling individual environmental obligations require not only behavioural alignment but also a shift in awareness, preferences, and understanding of one's relationship with the environment. This is possible when goals sought by GN are honest and accessible to the nudgee.

This recommendation addresses the societal critique of GN. Acting under wrong information or inaccessible behaviour goals takes away the opportunity to democratically engage nudgees with the issue of climate change (Button, 2018; Ewert, 2020). This also affects other aspects of the nudgee's life, such as increased susceptibility to Consumptive Nudging, discussed in the following subsection.

7.3. Minimising the scope of Consumptive Nudging:

As briefly outlined in Chapter 1, Consumptive Nudging (CN) refer to interventions designed to increase consumptive behaviours[108] in ways that are not fully understood by those being influenced (Braganza & Kapeller, 2025). It is driven by private, self-interested actors who leverage the behavioural features of nudgees for profit. Braganza and Kapeller (2025) argue that CN should be regulated to make GN effective[109].

CN are a concern for GN's effectiveness because private actors, motivated by profit rather than environmental benefits, are involved in the context in which GN takes place

[108] Consumptive behaviour refers to excessive purchasing behaviours that are driven by wants rather than by needs. Such behaviours are reflected in increased shopping habits, being influenced through social media platforms, and buying luxury goods to signal social status. Braganza and Kapeller's (2025) reference to consumptive behaviour is the kind of conspicuous consumption that has negative impact on environment. For example, the activity of driving a 'gas-guzzling SUV' discussed before.

[109] Their use of the term 'nudge' denotes behavioural influence without signifying beneficence. They point out that even Thaler and Sunstein (2008) contrary to widespread understanding of nudging emphasise that nudging can be used to increase sales (p.06)). Such commercial use of behavioural methods can undermine genuine GN.

(Braganza & Kapeller, 2025; Chater & Loewenstein, 2023; Gumbert, 2019a). Sunstein and Reisch (2014) also note that GN does not occur in a neutral, non-political context. It takes place in a system marked by market actors vigorously arguing for their own benefits. Braganza and Kapeller (2025) remark, "…this literature largely ignores what might be described as pervasive "anti-green" nudging by private actors" (p.03).

CN can undermine GN's effectiveness. The problem with CN is not a critique of GN; it concerns the circumstances within which GN and similar strategies promoting pro-environmental behaviour operate. Ivankovic and Engelen (2019) suggest that any governmentally implemented GN will be a drop in the ocean of privately motivated nudging strategies, and that any design of nudges should be informed by this consideration. Chater and Loewenstein (2023) observe that private actors use CN to resist systemic changes that threaten their interests, while framing climate change as an issue of individual responsibility. Button (2018) also remarks in this context,

> A policy focus on individual behavior modification fits very well with widespread corporate practices in the economy, and it is, for similar reasons, an effective way of avoiding the development of political consciousness and a shared sense of political responsibility for and efficacy in relation to societal challenges like climate change. (2018, p. 1040).

Exposing nudgees to corporate preferences negatively influences the engagement of nudgees with the issue of climate change, and/or depoliticises them as consumers. To address this societal critique, policymakers need to design GN interventions in a way that can increase the awareness of nudgees towards genuine behavioural changes.

Some ways to achieve this are by combining GN with other participatory approaches towards governance. Effective framing of narratives about climate change by local governments and the media, grounded in local values and norms, is one way to introduce

behavioural change (Romsdahl et al., 2018). Reforming media systems, strategising communications, framing debates, and incentivising pro-environmental behaviour are some of the strategies that can be combined with nudges.

On a similar note, boosting the competencies of the nudgee, incorporating reflection alongside the use of defaults or personalised nudging sought to reform the communication of the climate change agenda through behavioural insights (Banerjee & John, 2021; DellaValle & Sareen, 2020). Such strategies attempt to bring genuine change in behaviour, while ensuring the nudgee's active engagement with the issue of climate change.

Thus, the idea is that for GN to achieve their full potential, nudgees need to be incorporated as democratic participants. Since private actors actively use behavioural change strategies through CN, a lack of bottom-up citizen engagement increases the susceptibility of nudgees to corporate preferences (Chater & Loewenstein, 2023). While Button (2018) advocates nudging through democratic spaces, which can be a type of choice architecture or nudge by itself, Lenzi (2019) proposes integrating deliberative strategies with nudging.

An alternative approach is to combine the practice of nudging with self-governance to engage nudgees as democratic participants, protect objective values, such as principled autonomy, prudence, and fairness, foster trust, and build a collaborative space for reflection that can identify and regulate CN. In what follows, this proposal will be sketched and not fully fleshed out. The thesis has so far focused on examining the ethics of GN. A proposal to integrate GN with self-governance would involve an extensive undertaking with further analysis, which lies beyond the scope of this work. Nonetheless, the proposal signals directions for future research and offers a way to integrate nudgees as democratic participants (which is desirable from an ESC perspective).

7.4. Behavioural Self-Governance and Green Nudges:

The proposal to introduce GN into democratic spaces through self-governance addresses both societal and cultural critiques of GN. When nudgees are involved in the design and conceptualisation of GN, they also bring the contextual knowledge about culture, local norms, and language. Nudgees functioning as nudgers in the practice of designing, implementing, and regulating GN can ensure that GN is culturally sensitive and relevant. This contributes to the effectiveness and public acceptability of GN.

When citizens participate in the process of ensuring which behaviours are suitable for protecting local ecosystems, deliberate about generalisable rules, and participate in their implementation, it provides legitimacy for GN as an environmental measure, thereby addressing the societal critique of nudging. This suggestion is echoed in Button's (2018) work, which calls for citizens to function as choice-architects and to develop civic virtues while implementing nudges. Button (2018) writes,

> ...citizens can be mobilized to exercise their local knowledge and political agency to help imagine and create these institutional conditions, as they are in cities throughout the world where community organizations, private businesses, and local governments have dedicated themselves to nurturing the growth of compassion in their cities through reiterated and practical forms of action. (p. 1042).

While Button (2018) emphasises the role of citizens as choice architects, Hukkinen (2016) challenges the distinction between nudging and democratic behaviour. He argues that the use of behavioural influences, like norms, self-image, and social image, is prevalent in democratic contexts (Hukkinen, 2016). These behavioural influences are integral to how we understand the availability of choices, our behaviour, and decisions in democratic spheres. Reflective decisions made by democratic agents, such as voting, are driven by norms, emotions, and language. From this view, being nudged is a common place experience, even

in contexts that are seen as distinct from behaviour change scenarios. It is worth acknowledging that such an interpretation of nudging goes beyond its conventional usage, and draws on a broader understanding of behavioural influence that involve different dimensions of democratic participation.

Noting these two points, which are approached differently but share common insights, we ought to reflect on GN strategies through democratic spaces such as public forums, participatory processes, or public assemblies. One way to do so is by engaging nudgees in self-governance to improve behavioural outcomes. This can build culture and norms for addressing climate change at the local level while legitimising them as a strategy for driving sustainable behavioural transitions.

Building on these suggestions, this section reimagines the role of nudgees: seeing them as both the target of behaviour change and the nudger. This possibility has been considered by Button (2018) and Lenzi (2019). This proposal challenges the conventional practice of GN, which requires external agencies to intervene when individuals produce negative externalities. It considers the possibility of Behavioural Self-Governance (BSG) and emphasises a key point: *the expected behavioural change need not emerge through a top-down approach. It is possible to nudge through a bottom-up approach, letting nudgees take responsibility for their behaviour while building their capacities to adapt to behaviour change.*

Self-governance for managing individual behaviour in resource consumption is the primary theme in Elinor Ostrom's (1999, 2005) scholarship. It is also an ongoing project by Sheila Foster and Christian Iaione (2019) on urban governance. The project introduces behavioural governance in urban contexts, in light of technological advancements (Foster & Iaione, 2019). Almost 100+ cities have been mapped across the world to simulate self-

governance in urban structures (www.commoning.city). The idea of BSG for GN is sketched by drawing insights from these.

The discussion is structured as follows. First, an overview of self-governance and Behavioural Self-Governance (BSG) is provided. Second, the potential stakeholders in BSG are identified, and the design principles for self-governing GN strategies are outlined using existing examples of self-governance projects.

7.4.1. Self-Governance:

The challenge of distributing and regulating environmental resources through state or market mechanisms has intrigued policymakers and scholars alike, who find that both state and market mechanisms are often insufficient to address climate change (Foster & Iaione, 2022; Gumbert, 2019a; Ostrom, 1999). To address these limitations, alternative governance models have been explored. These governance models place individuals and communities at the centre of decision-making.

One such model is self-governance, which is a polycentric system. A Polycentric system is a spontaneous, self-correcting, and ordered system where individuals are the basic unit. Tocqueville's (1945) initial ideas for dividing the government machinery into smaller units evolved into the concept of polycentric systems, further influenced by Adam Smith, Michael Polanyi, George Stigler, Vincent Ostrom and Elinor Ostrom (Aligica & Boettke, 2009). This kind of system is a multi-layered system of governance with a number of economic and political actors involved: there are public officials in government, market actors interested in profit, and groups of individuals concerned with collective governance and the stewardship of shared resources (Aligica & Boettke, 2009).

Polycentric governance systems show how co-designed interdependent rules of behaviour allow everyone concerned with resource consumption to achieve mutual benefits over an extended period[110]. It captures the idea that while the government plays a role in facilitating, supporting, and organising shared resources, it does not hold the knowledge of how to govern those resources best in a given context. To address this limitation, individuals concerned with those resources are introduced as participants in the governance process.

[110] Like any governance model, there are limits to using psychological mechanisms and social norms to encourage resource sharing. For instance, when resource conflicts involve political or economic interests, voluntary cooperation alone may not suffice. In these cases, self-governance requires institutional arrangements to drive collective cooperation.

Multiple studies have tested this model and shown that, under the right circumstances, voluntary cooperation and self-governance can address collective action problems (Foster & Iaione, 2022).

Ostrom (1999) argued that conventional theories of government regulation and market-regulated systems assume that users of an environmental resource are alienated from one another, do not communicate, and are unwilling to enter into agreements to manage their behaviour within the system[111]. This, however, is not the case: individuals organise themselves into groups to make decisions about Common Pool Resources (CPR) under certain conditions (Ostrom, 1999).

She demonstrated through extensive field research how communities self-organise[112] to solve the Common-Pool-Resource (CPR) management issues. These collective action issues concerned limited supplies of resources, such as agricultural land, timber, and fisheries. Fieldwork conducted across Bolivia, Kenya, India, Madagascar, Tanzania, Nepal, and other countries offered an alternative vision to resource management to Ostrom (1999), whose work influenced US public policy in the late nineties. These self-organised systems were not fully detached from government authorities; they were embedded within formal governance structures and recognised by authorities.

[111] The intellectual history of Ostrom's work on self-governance and its popularity in the mainstream Western contexts can be traced back to the 1960s, when public administration systems in metropolitan areas across the United States were failing (Aligica & Boetke, 2009). There were overlapping jurisdictions among government units, duplication of functions, a lack of standardised procedures for crisis management, and insufficient coordination among units to exchange information in a timely and orderly manner (Aligica & Boettke, 2009).

[112] Individuals choose to join self-organised groups when the incentives to subscribe to the new rules of self-governance outweigh the combined costs, such as time and effort for self-organisation, the short-term cost of adopting new rules, and the long-term cost of maintaining a new self-governance system (Ostrom, 1999). For instance, a forest user may opt into a self-governance group when the resulting benefits outweigh the time and effort required to adopt the new rules, the short-term costs of implementing new strategies (such as newer equipment), and the long-term costs of monitoring and maintenance (including infrastructure maintenance and personnel hiring).

Self-governance systems open public sectors to innovation and encourage co-production and collaborative decision-making to address resource management and governance[113] issues. Ostrom's approach is ideal for CPR (lakes, rivers, forests). However, it does not account for asymmetric power distribution between markets and governments, governments and citizens, markets and citizens, and even within heterogeneous groups of citizens. Heterogeneity implies that differences in social, cultural, and economic contexts among citizens drive different understandings of how resources must be shared, regulated, and managed in a given context (Ostrom, 1999).

Thus, it becomes difficult to stimulate a self-governing system using Ostrom's principles within urban societies, which are larger and more diverse, with multiple types of stakeholders exhibiting economic, political, and informational asymmetries. Moreover, it is important to note that, at the time self-governing systems emerged in Ostrom's scholarship, the use of digital platforms was not common. Thus, while her principles considered the possibility of monitoring and surveillance in her systems, the principles did not account for modern technological advancements and algorithmic efficiencies.

There are differences between the cities and traditional CPR as investigated by Ostrom (1999). Cities are more populated and heterogeneous than the systems studied for self-governance. As the size of a group increases, many variables, such as trust levels, willingness to volunteer, and the costs of monitoring and conflict-resolution mechanisms, are influenced[114]. Moreover, when groups with different interests are present, self-governance is more difficult to achieve. Cities also possess mixed resources (both man-made and natural) and comprise both public and private property.

[113] Ostrom (1999) has outlined eight design principles that are exemplified by a long-term self-governance model. As these principles are integral to the notion of self-governance as proposed by Ostrom, these principles are discussed in Appendix B.

[114] A running hypothesis is that with the eight design principles by Ostrom, medium sized groups have a higher probability of succeeding in self-governance as compared to large groups (Ostrom, 1999).

In a quest to bring Ostrom's (1999) theory of polycentric governance to the city, Foster and Iaoine (2019) ask how urban resources, such as water supply and the use of public commons like parks and roads, can be managed through self-governance. They observe that many disputes in urban areas concern how to allocate finite resources to heterogeneous users in the most efficient way [115] (Foster & Iaione, 2019). The authors pose the question,

> Recognizing that there are many tangible and intangible urban resources on which differently situated individuals and communities depend to meet a variety of human needs, what might it look like to bring more polycentric tools to govern the city, or parts of the city, as a "commons?" (Foster & Iaione, 2019, p 347)

This question considers the possibility of self-governance in a city or urban space where residents from diverse groups, with varying population sizes and power asymmetries living in technologically advancing contexts, collaborate to govern resources.

At least two features of self-governance systems inspire a proposal of BSG. First, the success of climate policies and interventions depends on widespread behaviour change. Evidence from self-governance models indicates that a locally designed set of norms and interventions is more effective at changing behaviours than top-down strategies. This inspires BSG because, just as rules and regulations gain more legitimacy when shaped by communities, GNs informed by self-governing models will be more effective and drive long-term behaviour change, which is desirable from ESC.

Second, involving nudgees as stakeholders in the design of nudges enhances the public acceptability and trust in GN. When individuals help shape the rules and norms about

[115] It is important to consider the potential of self-governance in urban areas because cities are where a major chunk of the world's population would reside in the next century or so (Foster & Iaione, 2018). With growing population and increasing patterns of migration, urban areas would expand; covering more space, building more infrastructure, and inviting more migration with increased economic opportunities. Noting the need of self-governance in urban areas, the difference between CPR and urban resources, a different set of design principles for self-governing urban areas is developed (Foster & Iaione, 2019).

behaviour in pursuit of collective goals, the resulting strategies are more likely to be effective. This is because these rules reflect shared community values and interests, which are essential to trustworthiness. Moreover, collaborative strategies are likely to gain long-term support. In short, strategies for behaviour change are more likely to gain public acceptance if they align with the interests and values of the communities they target, as noted in subsection 6.3.1.

BSG thus offers a framework for improving collective behavioural outcomes through co-designed nudges, boosts, incentive programs, and digital behavioural tools, each aimed at maximising positive ecosystemic outcomes.

Although Ostrom's approach inspires BSG, it does not require a strictly polycentric form of governance from the outset. The framework proposed here remains open to different forms of bottom-up approaches that support self-governance, such as shared governance or collaborative governance. A shared form of governance is typically concerned with smaller groups and is evident in the use of resources such as parks and green spaces in neighbourhoods (Foster & Iaione, 2022). In contrast, a collaborative form of governance brings together bigger groups than a neighbourhood, involving civic societies, municipalities, local business owners, and the government. The proposal for self-governance initiated here can transition from a shared form of governance to collaborative governance, and ultimately to a fully polycentric form of governance, in which multiple independent decision-making units collaborate and exchange knowledge[116].

[116] Similar forms of governance have shown promise in Bologna and Madrid in improving the consumption of resources in urban areas, particularly for participatory budgeting, land trusts and co-ownership. The common factor between all these projects is that citizens qua nudgees are involved in the process of managing local natural resources and urban resources. They stimulate self-governance through a bottom-up approach, engaging users of a resource in the rules governing those resources, while recognising the role of the state as a manager of the process.

The following discussions are inspired by Ostrom's (1999) work and draw from the design principles proposed by Foster and Iaione (2019, 2022) in their work on urban co-governance. They outline principles that illustrate how behaviour strategies, especially GN, can emerge as outcomes of self-governance. Drawing on Foster and Iaione's (2019) work on urban co-governance, the following two subsections identify the main principles and stakeholders involved in BSG.

7.4.2. Stakeholders in Behavioural Self-Governance:

The first step in thinking of behaviour change as an outcome of self-governance is to identify a *self-governance collective* or the potential parties involved in the self-governance process. BSG is a co-governance arrangement among (at least) four stakeholders: a nudgee-citizen, a Behavioural Intermediary (BI), state representative(s), and private representative(s). Together, these four stakeholders–nudgee-citizens, the state, private actors, and Behavioural Intermediaries (BI)–can contribute to conceptualising, co-designing, and co-creating GN and behavioural insights that can drive transformative shifts.

Nudgee-Citizens. A nudgee-citizen is different from a nudgee because a nudgee-citizen is an active part of a self-governing structure. These are engaged in designing the behavioural strategies, identifying pathways to introduce behaviour change, and are also positioned to communicate with the institutions. They play a crucial role in legitimising behavioural policies by providing them a democratic character through participation (as well as deliberation), addressing societal critiques. Moreover, the co-production of behavioural knowledge through nudgees also improves the behavioural outcomes, thus addressing cultural critiques.

A BSG structure should not only aim to drive behavioural changes on a day-to-day basis, but also improve a nudgee-citizen's civic capacities, like mutual respect, tolerance, and

reciprocity, to address climate change[117]. Cultivation of these capacities alongside climate awareness drives more meaningful behavioural change to address climate change. This is morally desirable from an ESC view, as it fosters prudence and raises awareness of the environmental implications of one's behaviour.

Behavioural Intermediaries (BI). As already noted in 7.1, there is a gap between nudgees and nudgers, with the latter lacking the cultural and contextual knowledge to use nudges effectively. To address this divide between the government and the citizens, *Behavioural Intermediaries (BI)* can serve as points of communication between the nudgee-citizens and the state. Along with addressing the difficulties policymakers face in overcoming the knowledge gap, a BI can play an important role in legitimising behavioural strategies in democratic societies.

A BI may be an agent within autonomous institutions emerging from local communities. They may constitute elected representative(s) from a community, trained to maintain an effective dialogue with the state and private actors, and responsible for resources and facilities, such as organising public forums, citizen assemblies, and community engagement events for nudgee-citizens. A BI can act as a mediator between stakeholders, playing a critical role in transferring bottom-up knowledge to nudgers or state representative(s). This suggestion is inspired by the report from the Citizens' Assembly on climate change in Denmark that identifies a need for an intermediary that connects citizens (nudgee-citizens, in this case) to the state.

[117] Developing civic virtues through co-governance proposals has shown promise. For instance, Denmark established its first national Citizens' Assembly on climate change in December 2019. The resulting report emphasises the need to help citizens learn how and where to reduce their environmental impact. It discusses the economic aspects of behaviour change and the profound effect on life quality, longevity, and civic virtues. This suggestion emphasises the need for one to continue engaging in environmentally benign behaviour, even in the absence of nudging and behavioural regulations.

A BI as a stakeholder is crucial for two reasons. First, (BI) understands behavioural norms and local dynamics. A BI can facilitate participatory design processes and translate local knowledge and insights for states and vice versa, ultimately informing behaviour policy design. Second, a BI can act as a negotiator between the nudgee-citizens and the state (and/or corporations). For a BI, an outcome from this role would be gaining visibility, networking, and communicating their trustworthiness to the nudgee-citizens. For nudgers and the state, the result of having a BI would be increased perceptions of trustworthiness and fairness, better governance, and more effective behaviour change.

State Representatives. A state representative is responsible for initiating BSG in a community by bringing together nudgee-citizens, infrastructure, and resources. The Danish Citizen Assembly report mentions, "The politicians are responsible for making it easier for citizens and companies to change their behaviour and habits towards more climate-friendly decisions." (Borgertiget, 2021, p. 16). A behavioural policy expert appointed by the government can serve as a representative of the state, responsible for initiating and developing institutional infrastructure, such as policy innovation labs and collaborative programs. They can also be responsible for legal and financial mechanisms such as public-community partnerships and crowdfunding. Moreover, a state representative also oversees the running of technological infrastructure, such as digital applications and platforms, for public participation and deliberation.

Private Representatives. The private sector is an opportunity and a necessity for the BSG (Foster and Iaione, 2022). Private companies may take the initiative to facilitate activities that serve both commercial interests and the transition to a carbon-neutral society (Borgertiget, 2021). The stresses the importance of public-private partnerships as a critical part of BSG.

Considering the recommendation to minimise and regulate Consumption Nudging (CN) (6.2.2), it is important to engage private corporations in narratives around behavioural change agendas to ensure genuine behavioural change. Private corporations can play a role in providing resources such as labs for innovation and co-design to support BSG systems. In designing digital applications and platforms, corporations can blend profit and social motives through 'social corporate responsibility' practices, which are already adopted in the UK, the USA, and the EU (Foster & Iaione, 2018). For private corporations, self-governance can be an opportunity to build trust and acquire cooperation from nudgee-citizens who are also their consumers. Table 1 below summarises the role of these stakeholders.

Table 1: Stakeholders in Behavioural Self-Governance:

Stakeholder	Role	Contribution	Systemic Outcomes
Nudgee-Citizens	Individuals engaging in self-governance of GN.	Knowledge co-production, co-designing behavioural policies by participating in democratic processes.	Better behavioural outcomes, democratic participation.
Behavioural Intermediaries	Representatives who carry a greater responsibility of organising dialogue between the stakeholders.	Mediation between stakeholders, knowledge transfer about cultural intricacies, facilitation of resources and	Visibility, networking, communication of trustworthiness.

		capacity building programs.	
State-representatives	Government appointed agents who directly act as enablers for self-governance.	Resources, capacity building, and legitimising self-governance.	Building trust with stakeholders, preferential knowledge to improve the performance of BPP tools.
Private representatives	Representatives from private sectors and corporations in a local context responsible for maintaining Corporate Social Responsibility.	Funding acquisition, Capacity building.	Building trust with stakeholders.

7.4.3. Design Principles of Behavioural Self-Governance:

The rest of this subsection outlines the main design principles for BSG, based on Foster and Iaione's (2022) proposal of co-governance in urban areas. These principles also take into account suggestions from the Danish Citizens' Assembly, the European Union's PEAKApp project, Decidim.Barcelona, and the Amsterdam Smart City Project, amongst others. These design principles illustrate how BSG can be effectively stimulated through the four stakeholders.

The Enabling State. One of Ostrom's design principles for sustainable governance requires the state to recognise a community's right to govern a resource through collectively established rules and norms. Moreover, in a bottom-up approach, the state acts as an enabler by supporting the establishment and running of a co-governance system. On this basis, in BSG, the state must recognise the nudgee-citizens' ability and rights to make rules that shape choice environments and other interventions. To support BSG, three features of an enabling state are notable.

First, the state is the catalyst for self-governance (Foster & Iaione, 2022). One of the state's significant responsibilities is to reform existing institutional structures to introduce BSG (and bottom-up engagement) by nudgee-citizens. Second, the state is the network organiser. The state connects different networks, such as private corporations seeking avenues for social change, nudgee-citizens, and autonomous institutions. The state acknowledges their partnerships while regulating the distribution of resources.

Third, as certain actors (data-driven companies or big-data barons[118]) can unfairly exert behavioural influence over nudgee-citizens, the state can regulate power asymmetries within networks. The state also plays a role in providing a platform for bottom-up innovation through BIs. It has the responsibility of organising labs and participatory processes to allow nudgee-citizens to set their norms and expectations with collective behaviour.

Pooling. As Foster and Iaione (2022) note, "Our use of the term pooling instead describes the process of different sectors or actors combining their efforts to share resources, collaborate, and cooperate to create and steward urban goods, services, and infrastructure"

[118] Ranchordas (2019) points out 'big-data' barons as corporations capable of exerting strong behavioural influence on individuals because they possess data about one's preferences, background, and engagement obtained from one's digital footprint.

(p.45). In the context of BSG, pooling can be understood as the combined, collaborative efforts of the four stakeholders to optimise environmental behavioural outcomes.

Pooling is necessary to build better relationships among stakeholders, not just to achieve environmental goals. Developing better relationships leads to long-term systemic outcomes. For the state, pooling offers a way to promote welfare and a better trust relationship with the nudgee-citizens. For the markets, it supports engagement with environmentally concerned (potential or existing) consumers. Simply put, this principle drives the collective urban networks to adopt a bottom-up approach, coalescing stakeholders around social and economic platforms.

Pooling can be illustrated through practical examples. First is in the use of digital platforms for nudging energy consumption. An example of this is PeakApp, which aims to drive long-term energy-saving behaviour in households through nudges to maintain consistent energy behaviour via continuous engagement on social media platforms. It is a community-based energy conservation program that aims to serve the EU's target for smart-meter roll-out to transition energy-efficient citizens and consumers. This app brings together nudgees, energy providers, and nudgers on a digital platform to drive sustainable behaviour shifts through collective effort.

Second, engagement opportunities to pool the stakeholders for urban governance. Decidim.Barcelona is a digital co-governance platform launched by the Barcelona City Council to promote participatory governance and citizen engagement in public decision-making. The platform's primary mission is to empower citizens to shape their city's future through transparent, inclusive, and collaborative governance. Extensive neighbourhood-level participatory processes in this project offer residents platforms to co-create and co-design strategies to address social challenges, including air and noise pollution. An example is the

establishment of ‘Monitoring Committees’, made up of neighbourhood residents and organisations that meet regularly to track the progress of their projects.

Within this project, other instances of pooling include thematic working groups and ‘Inform and Involve’ mechanisms that engage the local community with the technical work related to urban commons. Moreover, through such citizenship engagement, ideas like superblocks (*superilles*) have also emerged: designing streets with residents in mind rather than vehicles has led to the emergence of new behavioural norms among residents and tourists, including walking rather than driving, places to sit and rest, and collective maintenance of green spaces. Such urban co-governance projects provide a template for introducing the principle of pooling in BSG.

Experimentalism. The primary purpose of adopting the principle of experimentalism in BSG is to integrate empirical evidence, lived experiences, and observations to inform and improve GN and similar behavioural policy tools. According to Foster and Iaione (2022), "Experimentalism allows localised knowledge and diverse observation data to enrich the process for developing local policies and practices that support the urban commons” (p.205).

One way to do so is to design Urban living labs or Policy innovation labs, which are typically sponsored by the state (Foster & Iaione, 2018). Three features of experimentalism emphasised in their approach are also interesting for BSG: Evaluation, Adaptation, and Interaction (Foster & Iaione, 2018).

Evaluation involves using qualitative and quantitative tools to assess the effectiveness of commons-based policies and practices. One straightforward example is the smart meter discussed by Sunstein and Reisch (2014). These tools enable real-time monitoring and quantification of energy behaviour and provide accessible feedback to nudgee-citizens (Allcott, 2011; Sunstein & Reisch, 2014). The European Union’s “Third European Energy Liberalisation package” also recommends the use of smart grids (or meters) to evaluate

energy outcomes while accounting for the data-privacy concerns of the users (Schubert, 2017b; Sunstein & Reisch, 2014). Another example is smart shower meters that provide real-time feedback on one's energy and water consumption during showers. The experiment demonstrated that digitally enabled behavioural interventions can induce substantial energy conservation (Tiefenbeck et al., 2019).

Adaptability involves recognising and incorporating contextual sensitivity, including historical and social dynamics (Foster and Iaione, 2022). An iterative approach to behavioural policy design is essential for interventions to be relevant and effective. As Małecka (2021, 2023) has already noted, failing to adapt to the context can result in the backfiring of behavioural interventions. In this context, the heterogeneous effects of nudging strategies (boomerang effect, moral licensing) are a much-discussed issue in the literature (Hilton et al., 2018; Schubert, 2017a), which emphasises the need for adaptable approaches.

Interaction refers to the engagement of a network of actors—such as residents, institutions, and local businesses—in the co-creation of interventions (Foster and Iaione, 2022). For BSG, inspiration can be drawn from the Amsterdam Smart City Platform, where the chief science office collaborates with academics and PhD students working in similar fields to develop evidence-based policies and advance an interoperable methodological language that can inform algorithmic interventions (Foster & Iaione, 2022). Learning from this approach, the principle of experimentalism can feed bottom-up knowledge into BSG systems through interaction and collaboration.

Tech-Justice. Given the advancement of modern technological tools, tech justice is a critical aspect of BSG. It taps into the recognition, participation, and distribution of the information of nudgee-citizens in digital spaces. Recognising the risks associated with algorithmic nudging and privacy, this is an important consideration.

This principle is mainly important for minimising the scope of hypernudging, which has become commonplace[119]. Yeung (2017) highlights that private entities collect, store, and process data to create tech-driven choice environments (nudge systems) that shape the choices of nudgees. This phenomenon is termed 'hypernudging' because, unlike a traditional behavioural strategy, such as sending one-time feedback on energy behaviour, hypernudging is a part of a dynamic, connected, and continuously adapting system. Yeung (2017) and Ranchordas (2019) observe that corporate organisations play an influential role in designing digital infrastructure that shapes collective behaviour. It is a powerful way to change behaviours, but when being used to fulfil corporate interests, it is dangerous.

Ranchordas (2019) notes that private corporations are key stakeholders in the design of smart-city infrastructure. In the absence of good governance, there is a risk of compromising the data integrity of the nudgee, creating a power asymmetry between corporations exerting behaviour influences and nudgee-citizens[120] (Ranchordás, 2019; Yeung, 2017). Drawing from Chapters 6s and 4, tech-justice also attempts to contribute to Privacy Measures (6.3.1).

In addition to regulating CN-based practices, tech-justice in BSG calls for technology-powered platforms that enable residents to propose policies, vote, develop solutions, and build digital literacy. PeakApp, an example discussed before, serves as a working template of how this principle can be manifested in behavioural governance.

By installing smart meters in housing societies, the app provides real-time data on energy consumption, gamifies energy-saving habits, and introduces social engagement with energy behaviour. It is designed with minimal hardware requirements (a single smart meter),

[119] Hypernudging is especially concerning as a subset of Consumption Nudging. As private corporations attempt to nudge individuals to increase consumption, they may adopt hypernudging.
[120] This feeds into neoliberal concerns about nudging, where behavioural interventions risk becoming tools for market-driven goals rather than democratic empowerment, a point already taken up about Consumptive Nudging (CN).

ensuring equitable distribution of technological resources and making energy facilities accessible to lower-income households. Similarly, by enabling users to gamify their behavioural outcomes and share them on social media, the initiative recognises their participation and uses this data to enhance future nudges.

The app has been tested in countries like Latvia, Sweden, and Austria, particularly to improve behavioural outcomes in social housing settings. Although PeakApp is not an ideal example because it does not directly engage nudgee-citizens in the design of the behavioural interventions, it remains an interesting illustration of how tech justice is an essential aspect of BSG.

7.5. Closing Remarks:

The chapter outlined three procedural recommendations for integrating GN into democratic societies. The first recommendation is to focus on honest framing and the accessibility of behavioural goals. The second recommendation is to regulate Consumptive Nudging (CN) to make sure that GN is effective. The third recommendation is to integrate GN into self-governance models to legitimise their use as democratic tools and to leverage the sociocultural knowledge available to the nudgees. The final proposal, Behavioural Self-Governance (BSG), remains tentative. Rather than offering a fixed/working definition of BSG, the chapter has focused on sketching design principles for such a mode of governance.

As these recommendations are mainly procedural, their outcomes will unfold over time. Nonetheless, some immediate effects can be seen in reviews and feedback on GN practices, public reporting on environmental impact, and reports of citizen assemblies for the participatory design of interventions.

The long-term effects of these recommendations will include the integration of behaviour interventions as a legitimate governance strategy, improved levels of trust and

trustworthiness, enhanced institutional capabilities and adaptation, and cognitive shifts and long-term behavioural enhancement.

This transformative shift in behaviour will be marked by nudgee-citizens who possess stronger environmental preferences, are capable of deliberating autonomously about environmental issues, and democratically self-govern how they are nudged. As Braganza and Kapeller (2025) note, this can be a process of self-realisation and growth as moral agents for the nudgees, with the primary goal of aligning behaviour and habits with one's values, a goal desirable from both a virtue-ethical and an ESC perspective.

Chapter 8. Conclusion

Given that the future, for lack of a better term, appears precarious, this thesis attempts to find a pragmatic and ethical solution to this issue with GN as a central case study. This research results from taking a systemic lens towards a global catastrophic challenge. The primary research question addressed in this dissertation is whether GN is ethically permissible for addressing climate change and, if so, under what conditions. By locating GN as a system of nudgee and the nudger, functioning at the intersection of systems, such as, the ecological systems, social and political systems, and technological systems (to an extent), this research introduces a consequentialist basis for making decisions about good systems.

Broadly construed, what has emerged out of this endeavour is an apparatus for examining the ethics of environmental systems, which is a pragmatic approach for reviving an ecocentric line of thought. More specifically, the ethical analysis of GN grounds behavioural ethics within environmental ethics, providing a strong justification for sustainable behavioural transitions.

In what follows, I conclude this discussion in the following sequence. In the next part (8.1), I summarise this research through three main arcs (or principle themes). Following this, I outline the broad insights from this study (8.2) and reflect on the contributions and future directions (8.3). Next, I consider some of the limitations of this research (8.4). Finally, I conclude with a reflection about the process of writing this dissertation (8.5).

8.1. Overview of the work

The dissertation follows three arcs (or main themes) to examine the ethics of GN. The first arc examines climate change as a systemic moral challenge and develops *Environmental System Consequentialism* (ESC) as a framework for ethical reasoning in light of the climate crisis (Chapters 1-2). The second arc explores the ethics of GN: its effectiveness in shaping

behaviour, the ethical issues it raises, and shows how GN can be defended from ESC, laying out the objective values needed for GN to be ethically permissible from this view (Chapters 3-5). The third arc considers the role of trust and other conditions for GN to be effective and ethical, before concluding with the broader implications of this work (Chapters 6-8). These arcs are revisited below to succinctly organise the main arguments of this research.

This first arc frames climate change as a systemic moral challenge and illustrates an ethical theory for acting in reference to this collective crisis. Ecological sciences show that climate change harms not only individual species but also the relationships that hold the ecosystems together; ethical reasoning must account for these interconnections. Existing individualistic and anthropocentric views do not account for the systemic nature of the crisis. Even though ecocentric theories frame the issue of climate change as a systemic crisis, they do not provide a pragmatic and moral basis to act through rules (or maxims).

Addressing this gap, a new theoretical approach, *Environmental System Consequentialism* (ESC), integrates ecocentric views with the traditional consequentialist views. It does so by providing a 'Rule-based Objective list' consequentialist basis to assess whether actions are morally justified from a systemic perspective: *actions are judged by the rules that maximise good ecosystemic outcomes, and also by the objective values that are necessary for ecosystems to flourish and for maintaining consistency with other existing human-centric systems.*

ESC revives ecocentrism as an ethical approach to make decisions concerning ecosystems, providing a consequentialist and pragmatic basis to judge actions in light of the climate crisis. This theory is a work in progress and is subject to conceptual refinement for its universal application to other systems warranting an ethical analysis.

The second arc explored the ethics of GN, its effectiveness in shaping behaviour, and the ethical issues it raises. To do so, it examined whether GN can be effective in helping us

deal with the climate crisis by working with the psychological shortcomings. It shows how, in practice, GN can act similar to low-level moral enhancement tools to help us fulfil individual obligations despite evolutionary psychological traits that undermine our capabilities to deal with a collective crisis.

Next, exploring the ethical issues associated with GN, led to a systematic review of the literature. Considering the need to address climate change, the inevitable ethical trade-offs associated with GN were also taken into account. A new understanding of autonomy–principled autonomy–was claimed, which is connected to individual obligations. The need to protect fairness and privacy was also emphasised. In light of the discussion about GN's effectiveness and the associated ethical issues, I could adopt a consequentialist position. With GN argued to be both effective and raising ethical challenges, I then explored the reasoning for its permissibility from ESC.

To respond to the central question of this thesis, it was explored how GN can be justified from an ESC position and what values and rules ought to be upheld to maximise good consequences for the ecosystem through GN. Three rules, about behavioural outcomes, awareness in the nudgee, and cognitive shifts were considered. Objective values pertaining to the ecosystems and objective values for maintaining consistency with other systems were considered.

The third arc shows trust as a functional value, offering recommendations for how green nudging can be ethically and effectively integrated into political and economic systems. Trust is argued as a functional and fundamental condition that makes it likely for GN, as a system between nudgee and the nudger, to be ethically justified from ESC. Since I have been examining systems, I have come to realise that the relationships between the different parts of ecosystems are valuable in themselves, which has led me to question the fundamental feature that makes relationships better or morally justified.

This value is trust between two or more parties/people/institutions. Drawing on Onora O'Neill's account of institutional trust and on the ubiquity of trust, as informed by Annette Baier's scholarship, I then reimagined the rules and signifiers of a morally justifiable relationship of trust between two components of GN, the nudger and the nudgee.

Finally, the circumstantial factors were explored. These recommendations make it more likely for GN to work as a legitimate tool in addressing climate change and upholding trust between the nudger and the nudgee. These recommendations, as I have already pointed out, have been somewhat procedural and limited because of the scope of this dissertation, which essentially focuses on ethics of GN.

8.2. Broad Reflections

At the outset, this thesis focused on examining the ethics of GN. In the literature on ethics of GN, the reasoning for using GN is framed as 'third-party' harm or an 'externality'. However, these approaches seem to overlook the rich literature in environmental ethics that can provide further ethical justification to the use of GN. On one hand, we see scientific evidence suggesting behavioural transitions as key for dealing with the collective crisis, and on the other hand, the ethical justification for using behaviour changes for pro-environmental ends are provided without exploring its normative dimensions in light of climate change as a moral issue (Nielsen et al., 2021). Looking at GN through the lenses of environmental theories allows me to show which human-centric values we ought to consider (fairness), trade-off (paternalism), or reframe (principled autonomy), along with ecosystemic goods.

Broadly construed, bringing perspectives from environmental ethics into traditional consequentialist views allows me to frame a pragmatic ecocentric position. This creates an interesting dilemma where, despite connecting environmental ethics to behavioural ethics, I have to some extent still upheld human-centric objective values of principled autonomy, prudence, and fairness. A potential objection could be that considering human systems as

valuable in outlining the ESC principles ignores the kind of 'oneness' that Naess (2005) argues for. Yet from another perspective, we might accept a pluralistic position that different objective goods are intrinsically valuable. Ultimately, choosing between human-centric goods and ecosystemic goods depends on the primary system of interest. Adapting ESC into a universal theory of systems may provide directions regarding this and remains a work in progress.

Interestingly, taking a systemic approach may reconcile fields like system dynamics in engineering and ecological sciences with applied ethics. A potential objection could come from the critics of systems-based thinking, that a systemic view in ethics may be biased due to the choice of factors, such as values considered as intrinsic goods, or the rules that maximise good ecosystemic outcomes. Ultimately, the epistemological limitations remain towards what one understands to be good outcomes and intrinsically valuable. Irrespective, looking at ethical decision-making through a systemic view is interesting as it enables us to map complex multidimensional issues like climate change into an ethical theory.

Lastly, without overclaiming what behavioural transitions can alone achieve without their integration into institutional systems, this dissertation highlights reasons and justifications to integrate GN leading to behavioural changes as a systemic strategy focusing on s-frames rather than only i-frames, a need also emphasised by Chater and Lowenstein (2023). ESC, the theoretical framework developed here, is aimed at providing pathways for integration of GN as such a strategy. Finally, it is important to clarify that rather than replacing systemic reforms, GN is aimed at complementing the myriad economic and political reforms that can help us deal with the collective action issue.

8.3. Contribution and Future Directions

First, a new consequentialist position is formulated. It adapts and modifies Hiller's (2011, 2013) version of system consequentialism. The thesis introduces ESC as a

consequentialist position for examining moral outcomes in the context of climate change. This ethical theory can be a fruitful avenue for examining the ethics of other contemporary issues.

Second, this thesis systematically reviews the ethical issues associated with GN and demonstrates that human moral psychology can impede morally desirable behavioural outcomes. While the idea that human moral psychology restricts moral actions has been explored in the transhumanist debate, this discussion extends the analysis to the realms of behavioural ethics and environmental ethics.

Third, it frames the future of the planet and societies in a new light. Any ethical examination ought to consider the complexities underlying a problem by examining how different systems interact and influence each other, as well as the potential trade-offs between values and outcomes that are required. This has the potential to mark a moral shift in the way we approach collective action issues. This position aligns with the radical shift occurring in the world, such as the climate crisis, the emergence of AGI, and locates it within a systemic view to better understand how we can better approach modern issues.

Taken together, these contributions open future opportunities for research in this area, particularly for an interdisciplinary approach in applied ethics. For GN, two research nodes are promising. First, the combined use of GN with the upcoming and advancing algorithmic processes is a relevant area of research. The discussions in this thesis do not yet consider the ethical issues beyond privacy. However, if large language models are deployed as nudgers in the future, an additional subset of the ethical problems pertaining to Artificial General Intelligence (AGI) should be considered. This is noted in the footnote (75) of this thesis.

The second direction/project of interest is the legitimisation of the use of GN and integrating them within self-governance models. Chapter 7 draws inspiration from real-world examples, such as, PeakApp, Co-Cities indicating how this ethical framework may inform

practice. Moreover, tying these examples to GN demonstrates how the use of behavioural insights can be embedded ethically into governance practices, addressing concerns regarding participation, fairness, and a balance between autonomy and collective goals. This can be integrated into an extensive interdisciplinary project that involves collaboration between the emerging field of Behavioural Public Policy (BPP), applied ethics, law, and social sciences.

Moreover, a research project structured around these two areas is a promising avenue. It appeals to an endeavour that combines the use of nudging and BPP developments within democratic practices, while mapping the limitations and potential of AGI in driving the process of accomplishing sustainable behavioural transitions..

8.4. Limitations

While the thesis has conducted a comprehensive ethical examination of GN, ranging from its need and effectiveness to ethical issues, system consequentialist defence, and conditions for justification, there remain some limitations to this study.

First, I have not fully developed an account of GN as a 'low-level moral enhancement tool' in Chapter 3. I have pointed out ways in which GN functions analogously to moral enhancement strategies, as well as how they differ from strategies traditionally considered as forms of moral enhancement, especially those targeting neurological and biochemical processes. My illustration of GN in light of moral enhancement theories does not commit me to accepting all kinds of moral enhancement as an ethically justified pursuit. This is because I use ESC as the main ethical theory informing the ethics of GN. There is an extensive debate on transhumanism and the need for moral enhancement in light of the collective action issues. My reasons for supporting GN are distinct from the transhumanist reasons. I have engaged with the moral enhancement debates only to the extent of showing why the reasoning of GN being effective is parallel with the efficacy of moral enhancement strategies.

Second, not all ethical issues and themes that were discussed in the literature review (Chapter 4) were addressed in this study. For instance, the sustainability goals of GN and the role of state and markets in using GN, are not directly addressed or developed in the thesis. I have examined GN as a strategy to enhance ecosystemic outcomes. While the pursuit of sustainability goals is not possible without engaging with the questions of how the systems between state/citizen, state/market, and citizen/market interact, these discussions tap into themes of political economy and social philosophy. Although an ethical theory of systems may explain the normative ideals of these relationships, it is a project for some other time.

Third, an objection could be that ESC assumes that certain objective goods, such as trust, fairness, and prudence, are necessary for maximising good outcomes. This criticism can be acknowledged in the same way as the objective list theories. For instance, an objection to the objective list theories is that values can be incommensurate with one another, or how do we prioritise one value over another when they conflict with each other (Mulgan, 2014). A potential response could be that an ESC or 'rule-based objective list system consequentialism' can resolve the problem of incommensurability between two values, or conflict between two objective values, by prioritising the objective values that uphold rules that maximise the long-term good outcomes for a system. Nonetheless, a fully fleshed-out version of an ESC theory is a work in progress.

Fourth, a previously considered objection is that the signifiers of trustworthiness outlined in Chapter 6 are not exhaustive. Even then, the two rules about morally justified trust are foundational for fostering trust between two parties. As such, the list of trustworthiness metrics or the discussion about nudgee's position to assess the evidence of trustworthiness is provisional and can be developed further through interdisciplinary collaborations.

Fifth, the proposal of Behavioural Self-Governance (BSG) has not been fully developed and is a tentative proposal. The goal of this research was to formulate and examine

the ethics of GN. To keep GN as the focus of the study, specific tangential but important research avenues have not been fully addressed in this context. This also involves developing and experimenting with the idea of BSG as a polycentric model of governance, a goal that is notable for future research.

8.5. Final Thoughts:

The way we understand the world is more often than not shaped by our academic training (amongst other environmental factors). In writing this dissertation, the biggest challenge was to understand how I should approach the issue of climate change as a moral crisis. In physics, universal laws are observed through systems that exist in the tiniest details of nature, such as the electromagnetic forces between particles on a plate, the gravitational pull between celestial bodies holding the universe together, or a falling apple from a tree. I found it justified and warranted to see climate change as a problem that is complex and can be defined through multitudes of systems that are at stake (and contribute) in this crisis. The result was an ethical examination that explored systems and subsequently, the rules, values, and connections necessary for the systems to flourish. To this end, GN served as the central case study.

Understanding this research question and responding to it has been a discovery narrative to a great extent: I started with an examination of the ethics of GN and subsequently understood my position as a philosopher who understands the moral challenges of our times through systems. Examining systems, the way I see it, unveils the complexities involved in moral decision-making. For instance, from this view, a good life can be achieved by preserving what leads to good consequences, but also by recognising that our actions do not occur in isolation; relationships, causal factors, and correlations are at stake.

GN, the main topic of this research, reflects my optimism about the possibility of sustainable behaviour transitions and the future of the planet. This optimism does not diminish the ethical concerns about GN that should be taken seriously. It is far easier to critique what can go wrong if we adapt to new technologies and developments that power GN. In contrast, I have tried to be hopeful in my discussions without losing sight of what is at stake if we do not act now, or of what could go wrong if we are not careful enough in the design, implementation, and regulation of GN and similar strategies.

References:

Albert, M. J. (2022). Beyond continuationism: climate change, economic growth, and the future of world (dis)order. *Cambridge Review of International Affairs*, *35*(6). https://doi.org/10.1080/09557571.2020.1825334

Aligica, P. D., & Boettke, P. J. (2009). Challenging institutional analysis and development: The Bloomington school. In *Challenging Institutional Analysis and Development: The Bloomington School*. Routledge. https://doi.org/10.4324/9780203876282

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095. https://doi.org/10.1016/j.jpubeco.2011.03.003

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. In *American Economic Review* (Vol. 104, Issue 10). https://doi.org/10.1257/aer.104.10.3003

Andersen, A., Karlsen, R., & Yu, W. (2018). Green transportation choices with IoT and smart nudging. In: Maheswaran, M., Badidi, E. (eds.), Handbook of Smart Cities: Software Services and Cyber Infrastructure (pp. 254–276), Springer. https://doi.org/10.1007/978-3-319-97271-8_13

Appleton, J. (2008). Reflections on Mark Lynas 'Six Degrees': Our Future on a Hotter Planet. Energy & Environment, 19(2), 371-376. https://doi-org.dcu.idm.oclc.org/10.1260/095830508783900807 (Original work published 2008)

Attfield, R. (2003). Environmental ethics: an overview for the twenty-first century / Robin Attfield. In *Environmental ethics: an overview for the twenty-first century*. Polity.

Attfield, R. (2012). Biocentrism and artificial life. *Environmental Values*, *21*(1). https://doi.org/10.3197/096327112X13225063228069

Baier, A. (1986). Trust and Antitrust. *Ethics*, *96*(2), 231–260

Banerjee, S., & John, P. (2021). *Nudge plus: incorporating reflection into behavioral public policy*. https://doi.org/10.1017/bpp.2021.6

Barandiaran, X. E., Monterde, A., & Rodríguez, L. (2024). Decidim, a technopolitical network for participatory democracy. In J. A. G. M. van Dijk (Ed.), Digital democracy: A critical overview (pp. 19–34). Springer. https://doi.org/10.1007/978-3-031-50784-7_1

Bennett, M. (2022). Judging Expert Trustworthiness: The Difference Between Believing and Following the Science. *Social Epistemology*, *36*(5). https://doi.org/10.1080/02691728.2022.2106459

Bentham, J. (1789). An introduction to the principles of morals and legislation. Oxford: Clarendon Press.

Biglieri, E. (2022). Beyond probability. *Dimensions of Uncertainty in Communication Engineering*, 203–251. https://doi.org/10.1016/B978-0-32-399275-6.00015-0

Bolderdijk, J., Steg, L., Geller, E. et al. Comparing the effectiveness of monetary versus moral motives in environmental campaigning. Nature Clim Change 3, 413–416 (2013a). https://doi.org/10.1038/nclimate1767

Bolderdijk, Jan & Gorsira, Madelijne & Keizer, Kees & Steg, Linda. (2013b). Values Determine the (In)Effectiveness of Informational Interventions in Promoting Pro-Environmental Behavior. PloS one. https://doi.org/10.1371/journal.pone.0083911

Bonini, N., & Dorigoni, A. (2024). Green Nudging: A Behavioral Approach to Environmental Policies. In P. Singh, K. Yadav, & S. Daga (Eds.), *Nudging Green: Behavioral Economics and Environmental Sustainability* (pp. 1–21). Springer. https://doi.org/10.1007/978-3-031-65972-0_1

Borgertinget (2021). *Borgertinget's recommendations on climate action: Citizens' Assembly on the climate crisis*. Danish Ministry of Climate, Energy and Utilities. https://www.kefm.dk/Media/637647201779892262/Borgertingets%20anbefalinger_ENG.pdf

Borneman, B., & Burger, P. (2019). Nudging to sustainability? Critical reflections on nudging from a theoretically informed sustainability perspective. In H. Straßheim & S. Beck (Eds.), *Handbook of behavioural change and public policy* (pp. 209–223). Edward Elgar Publishing Ltd. doi: 10.4337/9781785367854.00022

Bovens, L. (2009). The Ethics of Nudge. In Grüne-Yanoff & T.Hansson (Eds.), *Preference Change: Approaches from philosophy, economics psychology* (vol 42, pp. 207–219). Springer. Dordrecht. https://doi.org/10.1007/978-90-481-2593-7_10

Bovens, L. (2010). Nudges and Cultural Variance: a Note on Selinger and Whyte. *Knowledge, Technology & Policy*, *23*(3–4), 483–486. https://doi.org/10.1007/s12130-010-9128-2

Bowen, K. J., Patwardhan, A., Wang, Y., & Bialobrzeski, E. (2022). *The risk of cascading climate change shocks and stressors* (Science for Adaptation Policy Brief No. 5). The World Adaptation Science Programme (WASP) Secretariat, UNEP. https://apps.ipcc.ch/outreach/documents/specialevents/101/1665572139.pdf

Brady, W. J., & Crockett, M. J. (2024). Norm Psychology in the Digital Age: How Social Media Shapes the Cultural Evolution of Normativity. *Perspectives on psychological science : a journal of the Association for Psychological Science*, *19*(1), 62–64. https://doi.org/10.1177/17456916231187395

Braganza, O., & Kapeller, J. (2025). Reappraising consumption nudging—on liberty in the age of climate catastrophe. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-024-04246-0

Brezzi, M., González, S., Nguyen, D., & Prats, M. (2021). An updated OECD framework on drivers of trust in public institutions to meet current and future challenges. *OECD Working Papers on Public Governance*, *48*(48).

Broome, J. (2019). Against denialism. *The Monist*, 102(1), 110–129. https://doi.org/10.1093/monist/ony024

Brown, R. (2020). The citizen and trust in the (Trustworthy) state. *Public Policy and Administration*, *35*(4). https://doi.org/10.1177/0952076718811420

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv. https://doi.org/10.48550/arXiv.2303.12712

Button, M. E. (2018). Bounded rationality without bounded democracy: Nudges, democratic citizenship, and pathways for building civic capacity. *Perspectives on Politics*, *16*(4), 1034–1052. https://doi.org/10.1017/S1537592718002086

Byerly, H., Balmford, A., Ferraro, P. J., Hammond Wagner, C., Palchak, E., Polasky, S., Ricketts, T. H., Schwartz, A. J., & Fisher, B. (2018). Nudging pro-environmental behavior: evidence and opportunities. In *Ecology and the Environment* (Vol. 16, Issue 3).

Callicott, J. B., & Frodeman, R. (Eds.). (2009). Encyclopedia of environmental ethics and philosophy (Vols. 1–2). Macmillan Reference USA.

Cappa, F., Rosso, F., Giustiniano, L., & Porfiri, M. (2020). Nudging and citizen science: The effectiveness of feedback in energy-demand management. *Journal of Environmental Management*, *269*. https://doi.org/10.1016/j.jenvman.2020.110759

Capraro, V., Jagfeld, G., Klein, R. *et al.* Increasing altruistic and cooperative behaviour with simple moral nudges. *Sci Rep* 9, 11880 (2019). https://doi.org/10.1038/s41598-019-48094-4

Carlsson, F., Gravert, C., Johansson-Stenman, O., & Kurz, V. (2021). The use of green nudges as an environmental policy instrument. *Review of Environmental Economics and Policy*, *15*(2), 216–237. https://doi.org/10.1086/715524

Carter, L. (2025, February 12). Guide to investigating fossil fuels: Lobbying. Global Investigative Journalism Network. https://gijn.org/resource/guide-investigating-fossil-fuels-lobbying/

Chater, N., & Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, *46*. https://doi.org/10.1017/S0140525X22002023

Choe, E. K., Jung, J., Lee, B., & Fisher, K. (2013). Nudging people away from privacy-invasive mobile apps through visual framing. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8119 LNCS*(PART 3). https://doi.org/10.1007/978-3-642-40477-1_5

Christiansen, A. (2019). Rationality, Expected Utility Theory and the Precautionary Principle. Ethics, Policy & Environment, 22(1), 3–20. https://doi.org/10.1080/21550085.2019.1581413

City of Bologna. (2014). *Regulation on collaboration between citizens and the city for the care and regeneration of urban commons* (English trans. LabGov – LABoratory for the GOVernance of Commons). https://labgovcity.designforcommons.org/wp-content/uploads/sites/19/bolognaregulation.pdf

Congiu, L., & Moscati, I. (2022). A review of nudges: Definitions, justifications, effectiveness. *Journal of Economic Surveys*, *36*(1), 188–213. https://doi.org/10.1111/joes.12453

Copp, D. (2006). *The Oxford handbook of ethical theory*. Oxford University Press.

Council on Energy, Environment and Water. (2022). *Decoding changing monsoon rainfall patterns due to climate change in India*. https://www.ceew.in/publications/decoding-changing-monsoon-rainfall-patterns-due-to-climate-change-in-india

Courtenay-Hall, P., & Rogers, L. (2002). Gaps in Mind: Problems in environmental knowledge-behaviour modelling research. *Environmental Education Research*, *8*(3). https://doi.org/10.1080/13504620220145438

Daugbjerg, C., Smed, S., Andersen, L. M., & Schvartzman, Y. (2014). Improving Eco-labelling as an Environmental Policy Instrument: Knowledge, Trust and Organic Consumption. *Journal of Environmental Policy & Planning*, *16*(4), 559–575. https://doi.org/10.1080/1523908X.2013.879038

David, L. (2015). Social Identity Theory (Tajfel, Turner). *Social Psychology in Action*.

DeCew, J., & Roessler, B. (2023). Privacy. In E. M. Zalta & U. Nodelmann (Eds.), *The {Stanford} Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

DellaValle, N., & Sareen, S. (2020). Nudging and boosting for equity? Towards a behavioural economics of energy justice. *Energy Research and Social Science*, *68*. https://doi.org/10.1016/j.erss.2020.101589

Demarque, C., Charalambides, L., Hilton, D. J., & Waroquier, L. (2015). Nudging sustainable consumption: The use of descriptive norms to promote a minority behavior in a realistic online shopping environment. *Journal of Environmental Psychology*, *43*, 166–174. https://doi.org/10.1016/j.jenvp.2015.06.008

Des Jardins, J. R. (2000). Environmental ethics : an introduction to environmental philosophy / Joseph R. Des Jardins. In *Environmental ethics : an introduction to environmental philosophy* (3rd ed.). Wadswoth Thomson Learning.

Deshmane, A. (2023, August 31). *Inside Indian energy and mining giant Vedanta's campaign to weaken key environmental regulations*. Organized Crime and Corruption Reporting Project. https://www.occrp.org/en/investigation/inside-indian-energy-and-mining-giant-vedantas-campaign-to-weaken-key-environmental-regulations

DesRoches, T. C., Fischer, D., Silver, J., Arthur, P., Livernois, R., Crichlow, T., Hersch, G., Nagatsu, M., & Abbott, J. K. (2023). When is green nudging ethically permissible? *Current Opinion in Environmental Sustainability*, *60*, 101236. https://doi.org/10.1016/J.COSUST.2022.101236

DesRoches, C.T., Bartha, P., Mintz-Woo, K. *et al.* Climate nudging, catastrophes, and cost benefit analysis. *Euro Jnl Phil Sci* **16**, 39 (2026). https://doi.org/10.1007/s13194-026-00740-1

Dietz, A. (2019). Explaining the Paradox of Hedonism. *Australasian Journal of Philosophy*, *97*(3), 497–510. https://doi.org/10.1080/00048402.2018.1483409

Dobson, A. (2014). Improving Our Chances of Transition to Sustainability: The Role of Values and the Ethics of Solidarity and Sympathy. In *Addressing Tipping Points for a Precarious Future*. https://doi.org/10.5871/bacad/9780197265536.003.0026

Dold, M., & Stanton, A. (2021). I Choose for Myself, Therefore I Am: The Contours of Existentialist Behavioral Economics. *Erasmus Journal for Philosophy and Economics*, *14*(1). https://doi.org/10.23941/EJPE.V14I1.470

Drope, J., & Chapman, S. (2001). Tobacco industry efforts at discrediting scientific knowledge of environmental tobacco smoke: a review of internal industry documents. *Journal of epidemiology and community health*, *55*(8), 588–594. https://doi.org/10.1136/jech.55.8.588

Dworkin, Gerald, "Paternalism", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2020/entries/paternalism/>.

Ewert, B. (2020). Moving beyond the obsession with nudging individual behaviour: Towards a broader understanding of Behavioural Public Policy. *Public Policy and Administration*, *35*(3), 337–360. https://doi.org/10.1177/0952076719889090

Feizi, M., & Khatabiroudi, N. (2023). Social and environmental nudges and water usage: Evidence from a field experiment in Iran. *Water Resources and Economics*, *42*. https://doi.org/10.1016/j.wre.2023.100223

Feldman, S. (1988). Structure and Consistency in Public Opinion: the Role of Core Beliefs and Values. *American Journal of Political Science*, *32*(2). https://doi.org/10.2307/2111130

Fieser, J. (1992). The Compatibility of Eco-Centric Morality. *International Journal of Applied Philosophy*, *7*(1). https://doi.org/10.5840/ijap1992714

Forsberg, L., & Douglas, T. (2021). Moral enhancement. In *The Routledge Encyclopedia of Philosophy*. Routledge. https://doi.org/10.4324/9780415249126-L169-1

Foster, S. R., & Iaione, C. (2022). The city as a commons. In Co-cities: Innovative transitions toward just and self-sustaining communities (Chap. 3). MIT Press. https://doi.org/10.7551/mitpress/11702.003.0006

Foster, S., & Iaione, C. (2018). Ostrom in the City: Design Principles and Practices for the Urban Commons. In D. Cole, B. Hudson, & J. Rosenbloom (Eds.), *Routledge Handbook of the Study of the Commons* (1st ed., pp. 344–373). Routledge.

Frank L. E. (2020). What Do We Have to Lose? Offloading Through Moral Technologies: Moral Struggle and Progress. *Science and engineering ethics*, *26*(1), 369–385. https://doi.org/10.1007/s11948-019-00099-y

Fujii, S., & Taniguchi, A. (2006). Determinants of the effectiveness of travel feedback programs: A review of communicative mobility management measures for changing travel behaviour in Japan. *Transport Policy*, 13(5), 339–348. https://doi.org/10.1016/j.tranpol.2006.05.007

Gardiner, S. M. (2011). A Perfect Moral Storm: The Ethical Tragedy of Climate Change. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195379440.001.0001

Garvey, J. (2011). Climate Change and Causal Inefficacy: Why Go Green When It Makes No Difference? *Royal Institute of Philosophy Supplements*, *69*, 157–174. https://doi.org/10.1017/S1358246111000269

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3). https://doi.org/10.1037/0033-295x.103.3.59

Gigerenzer, G., Luan, S., & Liu, Y. (2019). Are we truly irrational and almost impossible to educate? Analyzing the scientific evidence behind libertine paternalism. *Acta Psychologica Sinica*, *51*(4), 395–406. https://doi.org/10.3724/SP.J.1041.2019.00395

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *The Journal of Consumer Research*, *35*(3), 472–482. https://doi.org/10.1086/586910

Goodin, R. E. (1993). Utility and the good. In P. Singer (Ed.), A companion to ethics (pp. 241–248). Blackwell Reference.

Goodwin, T. (2012). Why We Should Reject 'Nudge'. *Politics*, *32*(2), 85-92. https://doi.org/10.1111/j.1467-9256.2012.01430.x

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. In *Health Information and Libraries Journal* (Vol. 26, Issue 2, pp. 91–108). https://doi.org/10.1111/j.1471-1842.2009.00848.x

Gravert, C., & Shreedhar, G. (2022). Effective carbon taxes need green nudges. *Nature Climate Change*. https://doi.org/10.1038/s41558-022-01515-1

Grelle, S., Kuhn, S., Fuhrmann-Riebel, H., & Hofmann, W. (2024). The role of framing and effort in green nudging acceptance. Behavioural Public Policy, 8(4), 717–732. doi:10.1017/bpp.2024.8

Griffin, J. (1986). Well-being: Its meaning, measurement, and moral importance. Oxford: Clarendon Press.

Grimes, M. (2006). Organizing consent: The role of procedural fairness in political trust and compliance. *European Journal of Political Research*, *45*(2). https://doi.org/10.1111/j.1475-6765.2006.00299.x

Guala, F., & Mittone, L. (2015). A Political Justification of Nudging. *Review of Philosophy and Psychology*, *6*(3), 385–395. https://doi.org/10.1007/s13164-015-0241-8

Gumbert, T. (2019a). Anti-democratic tenets? Behavioural-economic imaginaries of a future food system. *Politics and Governance*, *7*(4), 94–104. https://doi.org/10.17645/pag.v7i4.2216

Gumbert, T. (2019b). Freedom, autonomy, and sustainable behaviours: The politics of designing consumer choice. In Isenhour C., Martiskainen M., & Middlemiss L. (Eds.), *Power and Politics in Sustainable Consumption Research and Practice* (pp. 107–123). Routledge. DOI:10.4324/9781315165509-5

Gupta, J., Chen, Y., McKay, D. I. A., Fezzigna, P., Gentile, G., Karg, A., van Vliet, L., Lade, S. J., & Jacobson, L. (2024). Applying earth system justice to phase out fossil fuels: Learning from the injustice of adopting 1.5 °C over 1 °C. International Environmental Agreements: Politics, Law and Economics, 24(1), 233–255. https://doi.org/10.1007/s10784-024-09628-y

Haines, A., & Patz, J. A. (2004). Health Effects of Climate Change. In *JAMA*. 291(1). https://doi.org/10.1001/jama.291.1.99

Hall, C. (2016). *Framing and Nudging for a Greener Future* (T. Gabrielson, C. Hall, J. M. Meyer, & D. Schlosberg, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199685271.013.15

Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, *7*(1), 155–174. https://doi.org/10.1017/S1867299X00005468

Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy. European Journal of Risk Regulation, 4(1), 3–28. doi:10.1017/S1867299X00002762

Hare, R. M. (1994). The structure of ethics and morals. In P. Singer (Ed.), Ethics (pp. 319–331). Oxford University Press.

Hardin, R. (2006). Trust. Polity.

Hardin, R. (2002). *Trust and trustworthiness*. Russell Sage Foundation.

Harmon-Jones, E., Harmon-Jones, C., & Levy, N. (2015). An Action-Based Model of Cognitive-Dissonance Processes. *Current Directions in Psychological Science*, *24*(3), 184-189. https://doi.org/10.1177/0963721414566449

Harvey, F. (2025, June 27). COP30: 10 things we learned from the Bonn climate talks. The Guardian. https://www.theguardian.com/environment/2025/jun/27/cop30-10-things-we-learned-from-bonn-climate-talks

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

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science*, *12*(6), 973–986. https://doi.org/10.1177/1745691617702496

Higgins, Tory E., (1989). Self-Discrepancy Theory: What Patterns of Self-belief Cause People to Suffer?. *Advances in Social Experimental Psychology, 22(1),* 93-136. https://doi.org/10.1016/S0065-2601(08)60306-8

Hiller, A. (2011). Climate Change and Individual Responsibility. *The Monist*, *94*(3). https://doi.org/10.5840/monist20119431

Hiller, A. (2013). System Consequentialism. In A. Hiller, R. Ilea, & L. Kahn (Eds.), *Consequentialism and environmental ethics* (pp. 85–111). Routledge.

Hiller, A. (2016). *Consequentialism in Environmental Ethics* (S. M. Gardiner & A. Thompson, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199941339.013.18

Hilton, D., Treich, N., Lazzara, G., & Tendil, P. (2018). Designing effective nudges that satisfy ethical constraints: the case of environmentally responsible behaviour. *Mind and Society*, *17*(1–2), 27–38. https://doi.org/10.1007/s11299-019-00201-8

Hoberg, N., & Strunz, S. (2018). When Individual Preferences Defy Sustainability — Can Merit Good Arguments Close the Gap? *Ecological Economics*, *143*, 286–293. https://doi.org/10.1016/j.ecolecon.2017.07.004

Holland, S., & Stocks, D. (2017). Trust and Its Role in the Medical Encounter. *Health Care Analysis*, *25*(3), 260–274. https://doi.org/10.1007/s10728-015-0293-z

Holton, R. (1994). Deciding to trust, coming to believe. Australasian Journal of Philosophy, 72(1), 63–76. https://doi.org/10.1080/0004840941234588

Hukkinen, J. I. (2016). Addressing the practical and ethical issues of nudging in environmental policy. *Environmental Values*, *25*(3), 329–351. https://doi.org/10.3197/096327116X14598445991501

Intergovernmental Panel on Climate Change. (2023). Climate change 2023: Synthesis report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC. https://www.ipcc.ch/report/sixth-assessment-report-cycle/

Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). (2019). Summary for policymakers of the global assessment report on biodiversity and ecosystem services. IPBES Secretariat. https://files.ipbes.net/ipbes-web-prod-public-files/inline/files/ipbes_global_assessment_report_summary_for_policymakers.pdf

Intergovernmental Panel on Climate Change. (2022). *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*. Cambridge University Press. https://doi.org/10.1017/9781009157940

Ivankovic, V., & Engelen, B. (2019). Nudging, Transparency, and Watchfulness. In *Social Theory and Practice* (Vol. 45, Issue 1). https://about.jstor.org/terms

Jones, K. (2012). Trustworthiness*. *Ethics*, *123*(1), 61–85. https://doi.org/10.1086/667838

Jones, K. (2019). Trust, distrust, and affective looping. *Philosophical Studies*, *176*(4). https://doi.org/10.1007/s11098-018-1221-5

Juengst, E., & Moseley, D. (2019). Human Enhancement. In E. N. Zalta (Ed.), *The {Stanford} Encyclopedia of Philosophy* (Summer 2019). Metaphysics Research Lab, Stanford University.

Kaaronen, R. O. (2018). Reframing Tacit Human–Nature Relations: An Inquiry into Process Philosophy and the Philosophy of Michael Polanyi. *Environmental Values*, *27*(2), 179-201. https://doi.org/10.3197/096327118X15162907484466

Kahneman, D., & Tversky, A. (1979). *Prospect Theory: An Analysis of Decision under Risk* (Vol. 47, Issue 2). https://about.jstor.org/terms

Kahneman, Daniel. (2011). *Thinking, Fast and Slow* . London: Penguin Books.

Kahrass (Knüppel), H., Borry, P., Gastmans, C., Ives, J., Graaf, R., Strech, D., & Mertz, M. (2021). *PRISMA-Ethics – Reporting Guideline for Systematic Reviews on Ethics Literature: development, explanations and examples*. https://doi.org/10.31219/osf.io/g5kfb

Kallbekken, S., Sælen, H., & Hermansen, E. (2012). 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

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

Kasperbauer, T. J. (2017). The permissibility of nudging for sustainable energy consumption. *Energy Policy*, *111*, 52–57. https://doi.org/10.1016/j.enpol.2017.09.015

Kavanagh, J., Carman, K., DeYoreo, M., Chandler, N., & Davis, L. (2020). The Drivers of Institutional Trust and Distrust: Exploring Components of Trustworthiness. In *The Drivers of Institutional Trust and Distrust: Exploring Components of Trustworthiness*. https://doi.org/10.7249/rra112-7

Kirby, P. (2024). Karl Polanyi in the transition to a low-carbon and biodiverse society. In M. Cangiani & C. Thomasberger (Eds.), The Routledge handbook on Karl Polanyi (pp. 361–372). Routledge. https://doi.org/10.4324/9781003336747

Knoll, L. (2019). Sustainable markets and the state: Taxation, cap-and-trade, pay-for-suecess, and nudging. *Historical Social Research*, *44*(1), 231–257. https://doi.org/10.12759/hsr.44.2019.1.231-257

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). https://doi.org/10.1080/13504620220145401

Krijnen, J. M. T., Tannenbaum, D., & Fox, C. R. (2017). Choice Architecture 2.0: Behavioral Policy as an Implicit Social Interaction. *Behavioral Science & Policy*, *3*(2). https://doi.org/10.1177/237946151700300202

Kuhn, M. (2025, September 22). *The impact of rising temperatures on worker productivity and health*. Climate Cosmos. https://climatecosmos.com/climate-society/the-impact-of-rising-temperatures-on-worker-productivity-and-health/

Kuusela, O. (2011). Consequentialism. In *Key terms in ethics* (pp. 18–22). Bloomsbury Publishing.

Kuyer, P., & Gordijn, B. (2023). Nudge in perspective: A systematic literature review on the ethical issues with nudging. Rationality and Society, 35(2), 191-230. https://doi.org/10.1177/10434631231155005

Lawrence, J., Blackett, P., & Cradock-Henry, N. A. (2020). Cascading climate change impacts and implications. *Climate Risk Management,* 29, Article 100234. https://doi.org/10.1016/j.crm.2020.100234

Leggett, W. (2014). The politics of behaviour change: nudge, neoliberalism and the state. *Policy & Politics*, *42*(1), 3-19. Retrieved Nov 15, 2025, from https://doi.org/10.1332/030557312X655576

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

Lembke, T.-B., Engelbrecht, N., Brendel, A. B., & Kolbe, L. M. (2020). To nudge or not to nudge: Ethical considerations of digital nudging based on its behavioral economics roots. In Proceedings of the 27th European Conference on Information Systems (ECIS 2019): Information Systems for a Sharing Society. Association for Information Systems. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087109333&partnerID=40&md5=cea5df33d5cf3a41d09ae19de1c74ae7

Lenzi, D. (2019). Deliberating about Climate Change: The Case for "Thinking and Nudging." *Moral Philosophy and Politics*, *6*(2), 313–336. https://doi.org/10.1515/mopp-2018-0034

Leopold, A. (1949). From A Sand County Almanac (1949). *A Sand County Almanac*.

Lepenies, R., & Małecka, 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

Li, J., Chiu, D. K. W., Ho, K. K. W., & So, S. (2024). The Use of Social Media in Sustainable Green Lifestyle Adoption: Social Media Influencers and Value Co-Creation. Sustainability, 16(3), 1133. https://doi.org/10.3390/su16031133

Loan, L. T. T., & Balanay, R. M. (2023). Towards reinforcing the waste separation at source for Vietnam's waste management: Insights from the Nudge Theory. *Environmental Challenges*, *10*. https://doi.org/10.1016/j.envc.2022.100660

Lorenz-Walther, B. A., Langen, N., Göbel, C., Engelmann, T., Bienge, K., Speck, M., & Teitscheid, P. (2019). What makes people leave LESS food? Testing effects of smaller

portions and information in a behavioral model. *Appetite*, 139, 127–144. https://doi.org/10.1016/j.appet.2019.03.026

Luo, Y., Hanson-Wright, B., Dowlatabadi, H. et al. How does personalized feedback on carbon emissions impact intended climate action?. *Environ Dev Sustain* 27, 3593–3607 (2025). https://doi.org/10.1007/s10668-023-04031-0

Lynas, M., Houlton, B. Z., & Perry, S. (2021). Greater than 99% consensus on human caused climate change in the peer-reviewed scientific literature. In *Environmental Research Letters* (Vol. 16, Issue 11). https://doi.org/10.1088/1748-9326/ac2966

MacIntyre, A. C. (1989). *A short history of ethics*. Routledge.

Małecka, M. (2021). Knowledge, behaviour, and policy: questioning the epistemic presuppositions of applying behavioural science in public policymaking. *Synthese*, *199*(1–2). https://doi.org/10.1007/s11229-021-03026-6

Małecka, M. (2023). Imaginary of Behavioral Governing. *Public Culture*, *35*(3). https://doi.org/10.1215/08992363-10742509

Marchiori, D. R., Adriaanse, M. A., & De Ridder, D. T. D. (2017). Unresolved questions in nudging research: Putting the psychology back in nudging. *Social and Personality Psychology Compass*, *11*(1). https://doi.org/10.1111/spc3.12297

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, *20*(3). https://doi.org/10.5465/amr.1995.9508080335

Mccrudden, C., & King, J. (2015). The Dark Side of Nudging: The Ethics, Political Economy, and Law of Libertarian Paternalism : Exploring the Legitimacy of Nudging. In A. Kemmerer, C. Möllers, M. Steinbeis, & G. Wagner (Eds.), *Choice Architecture in Democracies, Exploring the Legitimacy of Nudging* (pp. 75–140). Hart Nomo. DOI:10.5771/9783845263939-75

McLeod, C. (2023). Trust. In E. N. Zalta & U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy (Fall 2023 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entries/trust/

Mergoni, A., & De Witte, K. (2022). Policy evaluation and efficiency: a systematic literature review. *International Transactions in Operational Research*, *29*(3). https://doi.org/10.1111/itor.13012

Merz, J. J., Barnard, P., Rees, W. E., Smith, D., Maroni, M., Rhodes, C. J., Dederer, J. H., Bajaj, N., Joy, M. K., Wiedmann, T., & Sutherland, R. (2023). World scientists' warning: The behavioural crisis driving ecological overshoot. *Science progress*, *106*(3), 368504231201372. https://doi.org/10.1177/00368504231201372

Mill, J. S. (2006). Utilitarianism. In H. R. West (Ed.), The Blackwell guide to Mill's Utilitarianism (pp. 67–82). Oxford: Blackwell.

Miller, David, "Justice", *The Stanford Encyclopedia of Philosophy* (Spring 2025 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2025/entries/justice/>.
Mills, S. (2022). Finding the 'nudge' in hypernudge. *Technology in Society*, *71*, 102117. https://doi.org/10.1016/j.techsoc.2022.102117

Mills, S., & Sætra, H. S. (2022). The autonomous choice architect. *AI and Society*. https://doi.org/10.1007/s00146-022-01486-z

Mintz-Woo, K. (2025). Why We Should Support Carbon Prices. *Think*, *24*(70), 71–76. https://doi.org/10.1017/S1477175625000247

Mishra, A. K. (2012). Organizational Responses to Crisis: The Centrality of Trust. In *Trust in Organizations: Frontiers of Theory and Research*. https://doi.org/10.4135/9781452243610.n13

Mishra, V. (2025, February 26). *Somalia faces escalating crisis amid drought, conflict and price hikes*. UN News. https://news.un.org/en/story/2025/02/1160521

Moles, A.. (2015). Nudging for Liberals, *Social Theory and Practice,* 41(4), 644-667. https://doi.org/10.5840/soctheorpract20154143

Mulgan, T. (2014). III—Ethics for Possible Futures. *Proceedings of the Aristotelian Society*, *114*(1pt1), 57–73. https://doi.org/10.1111/j.1467-9264.2014.00364.x

Naess, A. (2005). The deep ecology movement: Some philosophical aspects. In A. Drengson & H. Glasser (Eds.), Selected works of Arne Naess (Vol. X, pp. 33–55). Springer.

Nagatsu, M. (2015). Social Nudges: Their Mechanisms and Justification. *Review of Philosophy and Psychology*, *6*(3). https://doi.org/10.1007/s13164-015-0245-4

Nefsky, J. (2021). Climate Change and Individual Obligations: A Dilemma for the Expected Utility Approach, and the Need for an Imperfect View. In *Philosophy and Climate Change*. https://doi.org/10.1093/oso/9780198796282.003.0010

Nefsky, J. (2019). Collective harm and the inefficacy problem. *Philosophy Compass*, 14(4), 12587. https://doi.org/10.1111/phc3.12587

Nicolao, B., Constantinos, H., & Michele, G. (2018). Green nudging. *Acta Psychologica Sinica*, *50*(8), 814. https://doi.org/10.3724/sp.j.1041.2018.00814

Nielsen, K. S., Bauer, J. M., Debnath, R., Emogor, C. A., Geiger, S. M., Ghai, S., Gwozdz, W., & Hahnel, U. J. J. (2024). Underestimation of personal carbon footprint inequality in four diverse countries. *Nature Climate Change*. https://doi.org/10.1038/s41558-024-02130-y

Nielsen, K. S., Clayton, S., Stern, P. C., Dietz, T., Capstick, S., & Whitmarsh, L. (2021a). How psychology can help limit climate change. *American Psychologist*, *76*(1), 130–144. https://doi.org/10.1037/amp0000624

Nielsen, K. S., Nicholas, K. A., Creutzig, F., Dietz, T., & Stern, P. C. (2021b). The role of high-socioeconomic-status people in locking in or rapidly reducing energy-driven greenhouse gas emissions. *Nature Energy*, *6*(11), 1011–1016. https://doi.org/10.1038/s41560-021-00900-y

Nolt, J. (2011). How Harmful Are the Average American's Greenhouse Gas Emissions?. *Ethics. Policy and Environment*. 3-10. 10.1080/21550085.2011.561584.

Nova, F. and L. Lades (2022), Nudges and Other Behavioural Public Policy Instruments to Encourage Environmentally Friendly Behaviour, University College Dublin, Dublin, https://publicpolicy.ie/environment/nudges-and-other-behavioural-public-policy-instrumentsto-encourage-environmentally-friendly-behaviour/ (accessed on 7 August 2024).

Nozick, R. (1974). Anarchy, state, and utopia. New York, NY: Basic Books.

Nykänen, E.-P., Hoppu, U., Löyttyniemi, E., & Sandell, M. (2022). Nudging Finnish Adults into Replacing Red Meat with Plant-Based Protein via Presenting Foods as Dish of the Day and Altering the Dish Sequence. *Nutrients*, *14*(19). https://doi.org/10.3390/nu14193973

O'Neill, O. (2002a). *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge University Press.

O'Neill, O. (2002b). Gaining autonomy and losing trust? In *Autonomy and Trust in Bioethics* (1st ed., pp. 1–27). Cambridge University Press.

O'Neill, O. (2002c). Autonomy, individuality, and consent. In *Autonomy and Trust in Bioethics* (1st ed., pp. 28–48). Cambridge University Press.

O'Neill, O. (2002d). Principled autonomy. *In Autonomy and Trust in Bioethics* (1st ed., pp. 73–95). Cambridge University Press.

O'Neill, O. (2002e). Principled autonomy and genetic technologies. In *Autonomy and Trust in Bioethics* (1st ed., pp. 96–117). Cambridge University Press.

O'Neill, O. (2002f). The quest for trustworthiness. In *Autonomy and Trust in Bioethics* (1st ed., pp. 118–140). Cambridge University Press.

O'Neill, O. (2018). Linking Trust to Trustworthiness. *International Journal of Philosophical Studies*, *26*(2), 293–300. https://doi.org/10.1080/09672559.2018.1454637

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

Oliver, A. (2013). From nudging to budging: Using behavioural economics to inform public sector policy. *Journal of Social Policy*, *42*(4). https://doi.org/10.1017/S0047279413000299

Oskari Kuusela, 'Consequentialism', in *Key Terms in Ethics* (London: Continuum, 2011), pp 18-21 (p.20).

Ostrom, E. (1999). *Self-governance and forest resources*. Center for International Forestry Research. https://doi.org/10.17528/cifor/000536

Ouvrard, B., & Stenger, A. (2019). Environmental regulation through nudges. In *Incentives and Environmental Policies: From Theory to Empirical Novelties*. https://doi.org/10.1002/9781119597490.ch3

Ouvrard, Benjamin & Spaeter, S.. (2015). Nudges as an alternative to economic market-based instruments in environmental regulation policies. In Daniel J.A. (Ed.), *Advances in Environmental Research* (Vol. 44, pp. 41–56). Nova Science Publishers. ISBN: 978-163483109-3;978-163483094-2

Padhy, S. K., Sarkar, S., Panigrahi, M., & Paul, S. (2015). Mental health effects of climate change. In *Indian Journal of Occupational and Environmental Medicine* (Vol. 19, Issue 1). https://doi.org/10.4103/0019-5278.156997

Parfit, D. (1984). What makes someone's life go best? In Reasons and persons (pp. 493–502). Oxford: Clarendon Press.

Partington, R. (2018, April 6). *Mark Carney warns climate change threat to financial system*. The Guardian. https://www.theguardian.com/business/2018/apr/06/mark-carney-warns-climate-change-threat-financial-system

PEAKapp Consortium. (2019). Personal energy administration kiosk application: An ICT ecosystem for energy savings through behavioural change, flexible tariffs and fun. https://www.peakapp.eu/publications/

Peña-López, I. (2017). *Decidim.barcelona, Spain: IT for Change case study*. The Institute of Development Studies and Partner Organisations. https://hdl.handle.net/20.500.12413/13284

Persson, I., & Savulescu, J. (2008). The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy*, *25*(3), 162–177. http://www.jstor.org/stable/24355031

Persson, I., & Savulescu, J. (2015). Summary of "Unfit for the Future." *Journal of Medical Ethics*, *41*(4), 338–339. http://www.jstor.org.dcu.idm.oclc.org/stable/44014059

Persson, I., & Savulescu, J. (2017a). The duty to be morally enhanced. Topoi, 38(1), 7–14. https://doi.org/10.1007/s11245-017-9475-7

Persson, I., & Savulescu, J. (2017b). Moral hard-wiring and moral enhancement. *Bioethics*, *31*(4). https://doi.org/10.1111/bioe.12314

Picard, J., & Banerjee, S. (2023, July 18). Unpacking the side effects of green nudges. SSRN. https://doi.org/10.2139/ssrn.4514424

Pichert, D., & Katsikopoulos V., K. (2011). Green Defaults: Information Presentation and Pro- environmental Behaviour. In *Heuristics: The Foundations of Adaptive Behavior* (pp. 713–724). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744282.001.0001

Placani, A. (2025). Individual responsibility for collective climate change harms. *Ethics, Policy & Environment*, 28(1), 79–94. https://doi.org/10.1080/21550085.2024.2347812

Population Reference Bureau. (2021, October 12). *Lifestyle choices affect U.S. impact on the environment*. https://www.prb.org/resources/lifestyle-choices-affect-u-s-impact-on-the-environment/

Potter, Nancy Nyquist, 2002, *How Can I be Trusted? A Virtue Theory of Trustworthiness*, Lanham, MD: Rowman & Littlefield.

Quélin, B., & Smadja, M. (2021). *Smart cities: The sustainable program of six leading cities* [Report]. HEC Paris. https://www.hec.edu/sites/default/files/documents/Amsterdam-Smartcities-the-sustainable-program-six-leading-cities-HECreport-2021.pdf

Raj, S. (2024, October 17). *In India, some doctors go on hunger strike to protest killing of colleagues. The New York Times.* https://www.nytimes.com/2024/10/17/world/asia/india-doctors-hunger-strike.html

Ranchordás, S. (2019). Nudging citizens through technology in smart cities. *International Review of Law, Computers & Technology*, *34*(3), 254–276. https://doi.org/10.1080/13600869.2019.1590928

Rawls, J. (1955). Two concepts of rules. *The Philosophical Review, 64*(1), 3–32.
Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy*, *6*(1). https://doi.org/10.1017/bpp.2020.5

Reinholdsson, T., Hedesström, M., Ejelöv, E., Hansla, A., Bergquist, M., Svenfelt, Å., & Nilsson, A. (2023). Nudging green food: The effects of a hedonic cue, menu position, a warm-glow cue, and a descriptive norm. *Journal of Consumer Behaviour*, *22*(3), 557–568. https://doi.org/10.1002/cb.2129

Riggs, W. (2019). The role of behavioral economics and social nudges in sustainable travel behavior. In *Transportation, Land Use, and Environmental Planning*. https://doi.org/10.1016/B978-0-12-815167-9.00014-1

Robertson, C. (2021). Reducing personal emissions in response to collective harm. Journal of Agricultural and Environmental Ethics, 34(14). https://doi.org/10.1007/s10806-021-09848-1

Rolston, H. (2011). The Future of Environmental Ethics. *Royal Institute of Philosophy Supplement*, *69*, 1–28. https://doi.org/10.1017/S135824611100018X

Romsdahl, R., Blue, G., & Kirilenko, A. (2018). Action on climate change requires deliberative framing at local governance level. *Climatic Change*, *149*(3–4). https://doi.org/10.1007/s10584-018-2240-0

Saghai. Y. (2013). Salvaging the concept of nudge. *Journal of Medical Ethics*, *39*(8), 496–497. https://doi.org/10.1136/medethics-2012-101060

Sajjad, M., & Khan, R. (2025, August 17). *At least 274 dead after devastating floods hit Pakistan*. Irish Independent. https://www.independent.ie/world-news/asia-pacific/at-least-274-dead-after-devastating-floods-hit-pakistan/a1397970445.html

Santos Silva, M. (2022). Nudging and Other Behaviourally Based Policies as Enablers for Environmental Sustainability. *Laws*, *11*(1), 9. https://doi.org/10.3390/laws11010009

Santos Silva, M. (2025). *A framework for good green nudging* (ELI Innovation Paper No. IP-2024-3). European Law Institute. https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/A_Framework_for_Good_Green_Nudging.pdf

Savulescu, J., & Bostrom, N. (Eds.). (2009). *Human enhancement*. Oxford University Press, Incorporated.

Savulescu, J., & Persson, I. (2012). Moral Enhancement, Freedom, and the God Machine. *Source: The Monist*, *95*(3), 399–421. https://www.jstor.org/stable/42751159

Schmidt, A. T., & Engelen, B. (2020). The ethics of nudging: An overview. *Philosophy Compass*, *15*(4). https://doi.org/10.1111/phc3.12658

Schubert, C. (2017a). Exploring the (behavioural) political economy of nudging. *Journal of Institutional Economics*, *13*(3), 499–522.

Schubert, C. (2017b). Green nudges: Do they work? Are they ethical? *Ecological Economics*, *132*, 329–342. https://doi.org/10.1016/j.ecolecon.2016.11.009

Schubert, S., Caviola, L., Savulescu, J., & Faber, N. S. (2024). *Temporal distance reduces ingroup favoritism*. https://doi.org/10.31234/osf.io/8k7bg

Scruton, R. (2002). A short history of modern philosophy (2nd ed.). Routledge Classics.

Selinger, E., & Whyte, K. (2011). Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture. *Sociology Compass*, *5*(10), 923–935. https://doi.org/10.1111/j.1751-9020.2011.00413.x

Selinger, E., & Whyte, K. P. (2010). Competence and Trust in Choice Architecture. *Knowledge, Technology & Policy*, *23*(3–4), 461–482. https://doi.org/10.1007/s12130-010-9127-3

Shahab, S., & Lades, L. K. (2021). Sludge and transaction costs. *Behavioural Public Policy* , 1–22. https://doi.org/10.1017/BPP.2021.12

Siipi, H., & Koi, P. (2022). The Ethics of Climate Nudges: Central Issues for Applying Choice Architecture Interventions to Climate Policy. *European Journal of Risk Regulation*, *13*(2), 1–18. https://doi.org/10.1017/err.2021.49

Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy & Public Affairs*, *1*(3), 229–243. https://about.jstor.org/terms

Singer, P. (1975). Animal Liberation: The Definitive Classic of the Animal Movement. *Preface to the Fortieth Anniversary Edition*.

Singer, P. (1981). *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Princeton University Press. http://www.jstor.org/stable/j.ctt7sg4n

Singer, P. (1993). About ethics. In Practical ethics (pp. 1–14). Cambridge University Press.

Singer, P. (2011). Climate Change. In P. Singer (Ed.), *Practical Ethics* (3rd ed., pp. 216–237). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511975950.010

Sinnott-Armstrong, W. (2005). It's not my fault: Global warming and individual moral obligations. In W. Sinnott-Armstrong & R. B. Howarth (Eds.), Perspectives on climate change: Science, economics, politics, ethics (Vol. 5, pp. 285–307). Emerald Group Publishing Limited. https://doi.org/10.1016/S1569-3740(05)05013-3

Sinnott-Armstrong, W. (2023). Consequentialism. In E. N. Zalta & U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy (Winter 2023 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/consequentialism/

Steel, D., Mintz-Woo, K., & DesRoches, C. T. (2023). Collapse, social tipping dynamics, and framing climate change. *Politics, Philosophy & Economics*, 23(3), 230-251. https://doi.org/10.1177/1470594X231196432 (Original work published 2024).

Steel, D., Vásquez-Fernández, A., Crookall, B., Cripps, R., DesRoches, C. T., & Mintz-Woo, K. (2026). Speed and Justice in a Renewable Energy Transition. *Ethics, Policy & Environment*, 1–19. https://doi.org/10.1080/21550085.2026.2614931

Steg, L. (2023). Psychology of Climate Change. *Annual Review Psychology*. 74:391-421. https://doi.org/10.1146/annurev-psych-032720-042905

Stern, N., Romani, M., Pierfederici, R., Braun, M., Barraclough, D., Lingeswaran, S., Weirich-Benet, E., & Niemann, N. (2025). Green and intelligent: the role of AI in the climate transition. *Npj Climate Action*, *4*(1), 56. https://doi.org/10.1038/s44168-025-00252-3

Sunstein, C. R., Reisch, L. A., & Kaiser, M. (2019). Trusting nudges? Lessons from an international survey. *Journal of European Public Policy*, *26*(10), 1417–1443. https://doi.org/10.1080/13501763.2018.1531912

Sunstein, C. & Reisch, L. (2014). Automatically Green: Behavioral Economics and Environmental Protection. The Harvard environmental law review: HELR. 38 (1). 127-158. https://doi.org/10.1111/kykl.12028

Sunstein, C. R. (2021). Green defaults can combat climate change. *Nature Human Behaviour*, *5*(5), 548–549. https://doi.org/10.1038/S41562-021-01071-2

Sunstein, C. R. (2023). Behavioral Biases, Choice Engines, and Paternalistic AI. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4539053

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian Paternalism Is Not an Oxymoron. The University of Chicago Law Review, 70(4), 1159–1202. https://doi.org/10.2307/1600573

Sustainability Directory. (2025). *Systemic values*. Lifestyle Sustainability Directory. https://lifestyle.sustainability-directory.com/term/systemic-values/
Szollosi, A., Wang-Ly, N., & Newell, B. R. (2025). Nudges for people who think. *Psychonomic bulletin & review*, 10.3758/s13423-024-02613-1. Advance online publication. https://doi.org/10.3758/s13423-024-02613-1

Tajfel, H., & Turner, J. C. (1985). The Social Identity Theory of Group Behaviour. *Psychology of Intergroup Relations, Vol2.*

Talbert, M. (2019). Moral responsibility. In E. M. Zalta & U. Nodelmann (Eds.), *Stanford Encyclopedia of Philosophy* (Fall 2024, Issue 5992). Metaphysics Research Lab, Stanford University.

Tannenbaum, D., Fox, C., & Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour*, 1, Article 0130. https://doi.org/10.1038/s41562-017-0130

Taylor, P. W. (1981). The ethics of respect for nature. Environmental Ethics, 3(3), 197–218. https://doi.org/10.5840/enviroethics19813321

Thaler, R. (2016). *Misbehaving – The Making of Behavioral Economics*.

Thaler, R. H., & Sunstein, C. R. (2008). Nudge improving decisions about health, wealth, and happiness. In *Nudge improving decisions about health, wealth, and happiness*. Yale Univ. P.

Thaler, R., & Sunstein, C. (2008). Saving the Planet. In *Nudge: Improving decisions about health, wealth, and happiness* (pp. 183–198). Penguin Books.

Tiefenbeck, V., Wörner, A., Schöb, S., Fleisch, E., & Staake, T. (2019). Publisher correction: Real-time feedback promotes energy conservation in the absence of volunteer selection bias and monetary incentives. *Nature Energy*, 4(10), Article 0484. https://doi.org/10.1038/s41560-019-0484-1

Tjeltveit, A. C. (2021, March 25). Ethics. In Oxford Research Encyclopedia of Psychology. Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.632

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Social cognitive and affective neuroscience*, *3*(2), 119–127. https://doi.org/10.1093/scan/nsn009

Tol, R. S. J. (2009). The economic effects of climate change. *Journal of Economic Perspectives*, *23*(2). https://doi.org/10.1257/jep.23.2.29

United Nations Climate Summit. (2023, April). *Comparing climate impacts at 1°C, 2°C, 3°C, and 4°C*. https://unclimatesummit.org/comparing-climate-impacts-at-1-5c-2c-3c-and-4c/

Van Gestel, L. C., Adriaanse, M. A., & De Ridder, D. T. D. (2021). Motivated by default—How nudges facilitate people to act in line with their motivation. *Motivation Science*. https://doi.org/10.1037/mot0000230

Varazzani, C., & Hubble, C. (2024). Four SINS in behavioural public policy. *Behavioural Public Policy*. https://doi.org/10.1017/bpp.2024.22

Viale, R. (2016). Bounded rationality and form of libertarian paternalism [Razionalità limitata e forme di paternalismo libertario]. *Sistemi Intelligenti*, *28*(1), 181–193. https://doi.org/10.1422/83843

Viale, R. (2018). The normative and descriptive weaknesses of behavioral economics-informed nudge: depowered paternalism and unjustified libertarianism. *Mind and Society*, *17*(1–2), 53–69. https://doi.org/10.1007/s11299-019-00207-2

Viale, R. Architecture of the mind and libertarian paternalism: Is the reversibility of system 1 nudges likely to happen*?*. *Mind Soc* 18, 143–166 (2019). https://doi.org/10.1007/s11299-019-00218-z

Vugts, A., Van Den Hoven, M., De Vet, E., & Verweij, M. (2020). How autonomy is understood in discussions on the ethics of nudging. *Behavioural Public Policy*, *4*(1), 108–123. https://doi.org/10.1017/bpp.2018.5

Wachner, J., Adriaanse, M., & De Ridder, D. (2020). The influence of nudge transparency on the experience of autonomy. Comprehensive Results in Social Psychology, 5(1–3), 49–63. https://doi.org/10.1080/23743603.2020.1808782

Walker, Margaret Urban (2006). Moral psychology. In Kittay Eva Feder & Martín Alcoff Linda, The Blackwell Guide to Feminist Philosophy. New York: Wiley-Blackwell. pp. 102-115.

Whitehouse, H. (2024). Copycat Culture. In *Inheritance:The Evolutionary Origins of the Modern World* (1st ed., Vol. 1, pp. 19–33). Penguin Random House Ireland.

Wilkinson, T. M. (2013). Nudging and manipulation. *Political Studies*, *61*(2), 341–355. https://doi.org/10.1111/j.1467-9248.2012.00974.x

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/2601248.2601268

World Meteorological Organization. (2025, January 10). *WMO confirms 2024 as warmest year on record at about 1.55°C above pre-industrial level*. https://wmo.int/media/news/wmo-confirms-2024-warmest-year-record-about-155degc-above-pre-industrial-level

Yeung, K. (2016). The Forms and Limits of Choice Architecture as a Tool of Government. *Law and Policy*, *38*(3), 186–210. https://doi.org/10.1111/lapo.12057

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information Communication and Society*, *20*(1). https://doi.org/10.1080/1369118X.2016.1186713

Appendix A: Systematic Methodology:

With the advent of evidence-based policymaking, the popularity of nudging has increased in sectors like health care, organ donation, finance, and environmental issues (Marchiori et al., 2017; Thaler & Sunstein, 2008). The literature review synthesises the ethical discussions, themes, and issues emerging in green nudges by harnessing the literature from this domain to make informed decisions (Grant & Booth, 2009). It attempts to systematically search and examine the published literature while adhering to PRISMA-Ethics guidelines (Kahrass (Knüppel) et al., 2021). Following the SALSA framework, the chapter aims for an exhaustive, comprehensive search of the literature on the ethics of green nudges (search), quality assessment (appraisal), narrative synthesis (synthesis), and analysis of potential gaps, uncertainty, and recommendations (analysis) (Grant & Booth, 2009).

This review takes an interdisciplinary approach, examining results from psychology, behavioural economics, political science, and ethics. The search results comprise research articles, review articles, and book chapters. The primary literature search was conducted using PRISMA-Ethics, a comprehensive and transparent method for systematic literature searches in ethics (Kahrass (Knüppel) et al., 2021). Additionally, a secondary search was conducted using Wohlin's Backward Snowballing method, which involves examining the reference list of relevant literature to identify useful results (Wohlin, 2014). The following subsections elaborate on the methodology used: Data Source, String Construction, Screening, and Results.

1. Data Sources and Search Strategies:

The search was conducted across five databases: Web of Science, Scopus, PhilPapers, JSTOR, and ScienceDirect. These databases provide access to peer-reviewed journals, articles, books, and book chapters. Search strings were used to access relevant results across the five databases using keywords in titles and abstracts. Advanced search options, wherever available at the databases, were utilised to make the searches comprehensive. Queries in advanced search engines used search strings constructed with keywords which were combined with Boolean operators such as AND, OR, and NOT to identify relevant results across the databases.

2. Construction of Search Strings:

The query in the search consisted of three strings. The first string involved the terms "nudge", nudges", and "nudging". The second search string consisted of terms such as, "ethic*", "moral*", "norm*", "libert*", combined with OR operator. The third search string consisted of terms related to climate change, "climat*", "anthropocen*", "green",environment*", put together with an OR operator. Asterisks were used to allow a greater variety of terms in the same context namely, environmentalism, anthropocentric, libertarian, normative, ethical, and so on. To obtain the search results having one or more terms from each of the three search strings, the three search strings were combined using the AND operator. A final search query was altered based on the efficiency of the search engines across the databases. For instance, PhilPapers was unable to handle a complex query consisting of a large number of keywords. A simpler query gave broadly relevant results on the ethics of green nudging on PhilPapers, and the results obtained on the database were sorted manually.

3. Use of Filters on Databases:

The search yielded a result of 430 articles across different journals. During the search, the language of the results was limited to 'English', and areas such as Engineering, Mathematics, and Accounting were excluded from the search. Keywords related to nudges were used in search engines. The search was conducted between 5th April 2022 to 14th April 2022. We used the available filters on various search engines ('Subject Area', 'Peer Reviewed', 'Search words appearing only in the Title and Abstract'). The exact search strings, and filters used in each of the databases are mentioned in Table 1 below.

Table 2: Details of Search Strings and Results:

Database Name	Search String and Filters used	Number of Results obtained
Web of Science	TS= ((“nudging” OR “nudge” OR “nudges”) AND (“ethic*” OR “moral*” OR “norm*” OR “libert*” OR “sustainab*”) AND (“climat*” OR “green” OR “anthropocen* “ OR “environment”)) Filters Used: Exclude Subject Areas: Health Care Sciences Services (2), Fisheries (2), Agriculture Dairy Animal Science (2), Hospitality Leisure Sports Tourism (10), Nutrition Dietetics (16), Food Science and Technology (11), Construction Building Technology (1), Neurosciences (3), Oceanography (3), All fields related to Engineering (7), Business (12). Limit Language to English.	181
Scopus	TITLE-ABS-KEY (“nudg*”) AND TITLE-ABS-KEY (ethic* OR moral* OR virtue* OR norm* OR “ libert*”) AND TITLE-ABS-KEY (“climat*” OR “anthropocen*” OR “green” OR “environment*” OR “sustainab*”) Keywords: Nudge, Nudges, Nudge Theory, Nudge Economics, Nudging. Filters Used: Limit to Language: English Exclude Subject Areas: Medicine (15), Accounting (10), Computer Science (7), Earth and Planetary Sciences (13), Nursing (13), Engineering (2),	17

	Decision Sciences (2), Mathematics (1), Chemical Engineering (1), Neuroscience (1), Pharmacology, Toxicology and Pharmaceutics (1), Veterinary (1).	
JSTOR	((ti: Nudge OR Nudges OR Nudging) AND (ethic* OR moral* OR libert* OR virtue *OR libert*)) AND (climate* OR environment* OR green OR sustainab* OR anthropocen*))	85
PhilPapers	(“nudg*” & (“climat*”|”green”|”anthrop*” | “environment*”))	40
Science Direct	(“nudge” OR “nudges” OR “nudging”) AND (“ethic” OR “moral”) AND (“green” OR “environment”) In title, abstract or keyword: Nudge, Nudging, Nudges	107

During the search duration, Science Direct and JSTOR search engines were unequipped to exclude subject areas, although they allowed researchers to include subject areas. Subject areas such as environmental management and psychology discuss ethical issues related to green nudges. To incorporate non-traditional discussions and themes on green nudges, any filters that restricted the search to include only specific subject areas were not applied on the databases. Limiting the search to certain research areas that are traditionally concerned with ethics could have led to the exclusion of relevant results on these databases. Due to this reason, subject filters that included only certain subject areas were not used.

The 430 search results exported in BibTeX or RIS format were uploaded on Zotero. In the Zotero library, 53 duplicate results were obtained. These automatically detected duplicates on Zotero were manually screened again. After this step, details of 377 results were exported into an Excel sheet for sorting the results categorically.

4. Screening:

377 results were screened initially. Firstly, the duplicate documents that were not detected by Zotero automatically were identified and removed manually in the screening process. Secondly, results that did not have any of the keywords “nudge”, “nudging”, or “nudges” in the title and/or abstract were excluded. This process led to the exclusion of 73 results in total.

304 papers containing keywords related to nudging were coded and screened based on their titles, and abstracts. All the results were sorted into the following categories:

Blue B1: These results do not specifically focus on the ethical issues arising from green nudges. Journal articles and book chapters categorised in B1 discuss the ethics of nudging under the assumption that all types of nudges, namely, nudges for savings, health, organ donation, and sustainable behaviour, raise similar ethical issues.

Light Green LG (Ethics of Green Nudges): The results focused on and contributed to the ethical discussions on green/environmental/climate/energy nudges only.

Peach P1: The results in this category discuss the ethics of nudges promoting good physical and mental health in individuals, savings, and organ donation.

Yellow Y1: Journal articles and book chapters under this category address themes of nudging in law and jurisprudence and their regulation and implementation by the government. Similar to the papers of category B1, results labelled as Y1 assume all kinds of nudges under the same category.

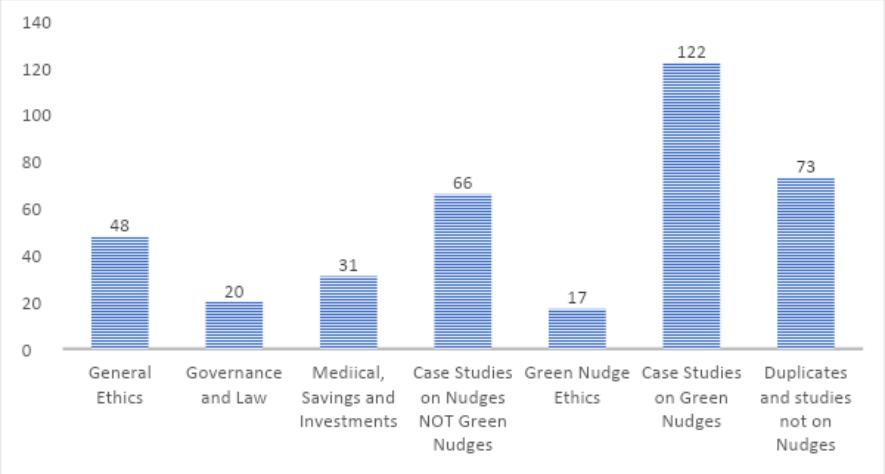
Pink P2: Papers in this category discuss case studies and empirical studies on all types of nudges except green nudges.

Dark Green DG: Results from the literature search that were either case studies or empirical studies on green nudges, were sorted into this category.

Table 3: Inclusion and Exclusion Criteria:

Step	Inclusion Criteria	Exclusion Criteria
Preliminary Screening	*Date Range:* Any *Journal Rank:* Any Keywords: Nudge, Nudges, Nudging	1. Articles not in English. 2. Newspaper articles, Conference proceedings, PhD theses. 3. Duplicate Results.
Title and Abstract Screening	Papers coded as LG (17). **Criteria:** Does the result exclusively examine the ethics of environmental/green/climate nudges? If yes, then the paper was coded as LG.	Papers coded: B1 (48), P1(31), Y1(20), P2(66), DG (122).

Graph 1: Distribution of results into different categories during the screening process:

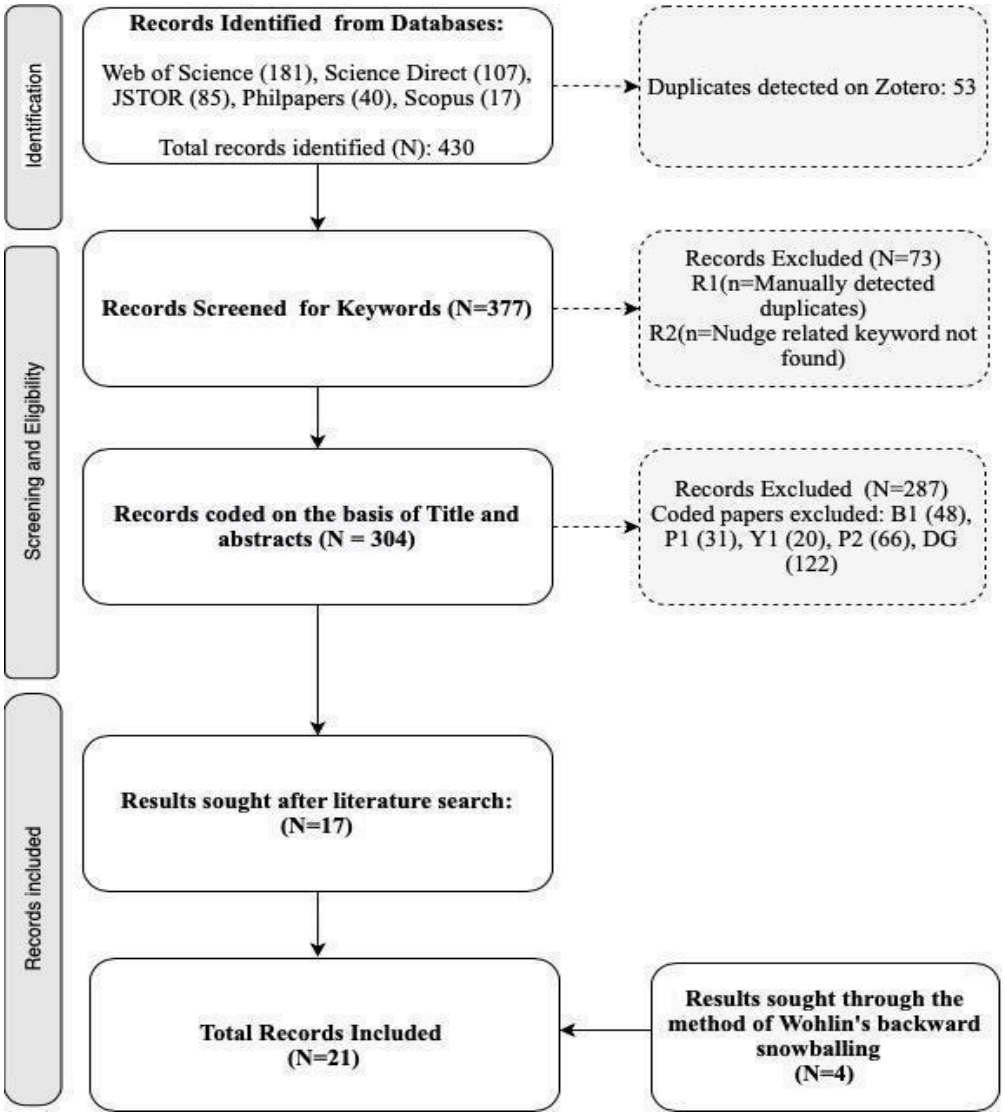


5. Results:

The screening process led to the exclusion of 289 results. In total, 17 results consisting of journal articles and book chapters that fit the criteria were selected for the literature review (Carlsson et al., 2021; DellaValle & Sareen, 2020; Dobson, 2014; Gumbert, 2019a; Hall, 2016; Hilton et al., 2018; Hoberg & Strunz, 2018; Kaaronen, 2018; Kasperbauer, 2017; Knoll, 2019; Lehner et al., 2016; Lenzi, 2019; Ouvrard & Stenger, 2019; Santos Silva, 2022; Schubert, 2017b; Siipi & Koi, 2022; R. Thaler & Sunstein, 2008).

Following this, Wohlin's backward snowballing was used to identify other relevant results (Wohlin, 2014). This was the secondary stage of the search. A process of Wohlin's backward snowballing involves going through the bibliography or reference list of the results to identify relevant literature. This step gave us four additional results consisting of book chapters and journal articles (Borneman & Burger, 2019; Gumbert, 2019b; Hukkinen, 2016; Ouvrard & Spaeter, 2015). Subsequently, 21 results were analysed in this literature review. A PRISMA flow diagram of the literature search is given below.

Figure 2: A four-step PRISMA-Flow diagram of the literature search



Appendix B: Design Principles by Ostrom (1999):

P1. Boundaries: There are rules governing who has the right to use a resource and the limits of that resource.

P2. Congruence: This principle has two parts. First, the rules governing the costs and benefits associated with resource usage are considered fair and legitimate by the users. Second, the rules related to time, place, and technology must be realistic and relevant to the users.

P3. Collective Choice Arrangements: Individuals affected by the system can participate in the operationalization of the system.

P4. Monitoring: The system has monitors responsible for auditing conditions and user behaviour.

P5. Graduated Sanctions: Sanctions exist for those who violate the rules of the system, depending on the seriousness of the violation and the context.

P6. Conflict Resolution Mechanisms: Different individuals within the group have low-cost access to conflict resolution mechanisms.

P7. Recognition: External authorities do not challenge the group's right to self-govern.

P8. Nested Enterprise: For CPR, which is part of a larger system, appropriation, provision, and conflict resolution occur through a multilayered enterprise.

Appendix C: Design Principles in Co-Cities Project (2019):

Ostrom's proposal of self-governance is applied to the cities to reimagine the design principles by Forster and Iaione (2019, 2022). Considering the set of differences (such as diverse population, politics, congestion, and socio-cultural complexity), five design principles for a city are proposed:

Principle 1 (P'1): Collective Governance: A city constitutes multiple stakeholders in governance where the community constitutes various actors and partners for collaboration, co-design, and coordination. Such stakeholders, such as citizens, public authorities, market agents, civil society organisations, and knowledge institutions, together form a loosely coupled system aimed at self-governance of resources in urban areas.

Principle 2 (P'2): Enabling State: The state is the enabler (usually public authorities) that facilitates self-governance.

Principle 3 (P'3): Social and Economic Pooling: Pooling refers to the activity of different stakeholders combining their knowledge, experience, and efforts to better govern the urban goods, infrastructure, and services. The transition of the traditional form of urban governance towards a more polycentric self-governing structure is marked by a process of social and economic pooling.

Principle 4 (P'4): Experimentalism: A place-based iterative approach for co-designing innovative strategies for self-governance.

Principle 5 (P'5): Tech-justice: Co-ownership and co-designing of digital and technological infrastructure for the creation of self-governance in urban contexts.