



Universitat Autònoma de Barcelona

**ADVERTIMENT.** L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  [http://cat.creativecommons.org/?page\\_id=184](http://cat.creativecommons.org/?page_id=184)

**ADVERTENCIA.** El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

**WARNING.** The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



# **The acceptability of nudges as public policy tools: a theoretical and empirical analysis**

**Júlia de Quintana Medina**

PhD thesis

Department of Sociology, Universitat Autònoma de Barcelona

2020

**Author:**

Júlia de Quintana Medina

PhD Programme in Sociology, Department of Sociology, Universitat Autònoma de Barcelona

**Supervisor:**

José Antonio Noguera Ferrer, Department of Sociology, Universitat Autònoma de Barcelona



# Acknowledgements

L'elaboració d'una tesi és un procés difícil i solitari i que m'hagués estat impossible sense el suport de moltes persones. Els vull donar les gràcies per haver contribuït d'una manera o altra a afrontar aquest procés acompanyada.

En primer lloc, vull donar les gràcies al meu director, el José Noguera. Gràcies José, per la paciència, la confiança i pels teus valuosos comentaris. Gràcies també per incloure'm en projectes que m'han permès formar-me com a sociòloga i per, des del primer moment, tractar-me com un igual i valorar la meua opinió. Que algú a qui admires professionalment t'inclogui i confi amb el teu criteri és un dels millors reconeixements possibles.

Treballar al GSADI ha estat un plaer, dono les gràcies a tots els meus companys, en especial a la Judith i l'Oriol. Gràcies Judith pel suport moral i acadèmic i per tots els consells personals i professionals. Gràcies Oriol per atendre i resoldre tots els meus dubtes de programació i sobretot per aguantar la meua xarrera i de tant en tant riure amb alguna de les meves bromes. També vull mostrar el meu agraïment al Departament de Sociologia i en especial a la Mònica i l'Elisabet per ajudar-me a navegar els tràmits administratius i ser sempre agradables amb mi.

En segon lloc, vull donar les gràcies a tots els meus professors i professores que, quan he estat alumne, han confiat amb les meves capacitats i m'han animat a seguir estudiant. Vull agrair especialment a Ferran Teixidor, el meu professor de filosofia i sociologia a batxillerat. Sense les seves classes dubto molt que hagués desenvolupat el mateix interès per les qüestions socials i segurament no hagués seguit el mateix camí. El doctorat és un espai que et permet treballar però formar-te i seguir sent alumne, que és en realitat el que a mi més m'agrada. Ha estat gràcies a la confiança dels meus professors i professores que he pogut seguir estudiant i seguir aprenent, així que el meu agraïment sincer per tots ells i elles.

En tercer lloc, la tesi no hauria estat possible sense la companyia i el suport dels que tinc a prop i que durant aquests anys m'han ajudat de moltes maneres.

A les meves amigues de sempre, a qui conec des de la guarderia i que són la millor constant a la meua vida. Gràcies Helena, Marta, Paula, Clara, Anna, Alba i Carmina per no preguntar massa i mostrar-me suport incondicional. Gràcies per no jutjar els moments en què he estat més distant i per distreure'm amb in comptables partides a la botifarra. Créixer juntes ha estat una sort, sou les millors.

A l'Ariadna i a la Núria, de les millors coses que m'emporto d'estudiar sociologia. Discutir amb vosaltres de temes polítics i vitals és sempre estimulant, em recorda totes les coses que ignoro i m'ajuda a ser millor sociòloga. Està a prop de dones fortes i intel·ligents com vosaltres és molt inspirador, i em sento molt afortunada de tenir-vos com amigues.

A tots els meus altres amics i amigues i totes les persones que en algun moment s'han interessat per com anava la tesi, per com estava jo i m'han ofert ajuda i suport. Gràcies.

A tota la gent que he conegut gràcies a en Marc i que m'han acompanyat també aquests anys. Gràcies a la seva família, a la Marta i en Joan i a la Núria en Roger i en Nau per acollir-me i tractar-me tan bé. Gràcies també al Raul i la Laura perquè són encantadors i perquè jugant al Catan m'han distret moltes tardes quan la tesi era més difícil de pair.

A la meva família, el meu suport més important i a qui dec tot el que sóc avui. Als meus avis, en Manel i la Lluïsa perquè tot el que faig els sembla bé i sempre recorden comentar-m'ho. Al meu pare, en Xavi. Gràcies papa per ensenyar-me què és gaudir de la feina acadèmica, i molt més important, per ensenyar-me que a la vida s'ha de ser empàtic, amable i honest. A la meva mare, la Mercè. Gràcies mama per cuidar-me, ensenyar-me a tenir veu i animar-me sempre a ser valenta i independent. També gràcies mama perquè és al segona tesi que pateixes i ho has tolerat força bé. Al meu germà, en Jaume, amb qui tot i ser diferents ho comparteixo tot. Gràcies Jaume per fer-me riure, per confiar amb mi i per recordar-me que no cal agafar-se les coses sempre tant seriosament. Espero, a partir d'ara, poder ser tan comprensiva i positiva com tu ets amb mi amb tots els teus projectes.

Finalment, tot això no hauria estat possible sense en Marc. Gràcies per la paciència, el suport, l'energia i per confiar més a mi que jo mateixa. La tesi és mig teva. No sé com donar-te les gràcies, només espero poder-t'ho tornar en el futur d'alguna manera.

## Summary

This thesis is a theoretical and empirical analysis of the acceptability of nudges as policy tools. The first part of the thesis discusses the characterisation of nudges as a policy instrument, paying attention to conceptual, ethical and theoretical issues. The thesis looks at the original notion of nudges, their theoretical and empirical bases and the main criticisms of both dimensions. It goes on to outline the problems with Thaler and Sunstein's notion of nudges and proposes an alternative understanding of nudges, advocating dropping the normative connotations of nudges, and focusing on their practical value as a policy tool. The second part includes three empirical studies on attitudes towards nudges: a nationally representative survey ( $n=617$ ,  $e=3.88$ ) of the public acceptability of nudges in Spain, and two experimental studies that explore the acceptability of two types of nudges, automatic nudges that are less noticeable, work with no deliberation and awareness, and reflective nudges that are noticeable, work by engaging deliberation and produce a more reflective response. The empirical part addresses some of the concerns discussed in the theoretical debate and emphasises that the discussion on the acceptability of nudges as policy tools should include considerations about whether people like nudges, which nudges they prefer and the contexts in which they accept their implementation. The results outline three main themes. Firstly, the public acceptability of nudges is high, people favour the use of different nudges applied to both pro-self and pro-social domains and the indication is that people do not categorically oppose nudges. Secondly, the type of nudge is important. People recognise differences between nudges and tend to prefer reflective nudges to automatic nudges. At the same time, the findings in the three studies indicate that people do not categorically oppose the implementation of automatic nudges and support their use in several cases. Thirdly, the acceptability of automatic nudges depends on the context in which they are applied; people support their use when they agree with the aim and when they affect behaviour in low-stake domains. By contrast, their use is not supported when they affect decisions in high-stake domains, such as decisions that involve sensitive or moral issues or decisions that involve economic losses. Overall, the thesis explores arguments in favour of and against nudges and offers ideas and conclusions to give an idea of how nudges could be useful and acceptable tools for policymaking.

# Contents

<b>Introduction.....</b>	<b>1</b>
1. Research objectives and questions.....	5
2. Outline of the thesis .....	5
<b>Chapter 1 The idea of nudge: origins, foundations and problems.....</b>	<b>8</b>
1. Introduction.....	8
2. The original nudge characterisation.....	9
2.1. Nudge and behavioural economics .....	10
2.2. Nudge and libertarian paternalism .....	12
2.3. The original nudge characterisation, five definitional properties. ....	14
3. Problems with the arguments that support the original nudge characterisation .....	14
3.1. Problems with the conception of rationality .....	16
3.2. Problems with the ways in which nudges steer choices.....	19
4. The need for a new understanding of nudges .....	20
<b>Chapter 2 The ethics of nudges .....</b>	<b>23</b>
1. Introduction.....	23
2. Objections to the aims of nudges.....	24
2.1. Nudges and wellbeing, as judged by individuals themselves? .....	25
2.2. Nudges, alternative aims, and justifications.....	29
3. Objections to the means of nudges .....	32
3.1. The rationality objection .....	34
3.2. The reflection objection .....	40
4. The moral acceptability of nudges: a combination of three aspects.....	65
4.1. The aim.....	66
4.2. The type of nudge.....	67
4.3. The context of choice.....	75
5. Concluding remarks.....	81
<b>Chapter 3 The use of behavioural insights in public policy: opportunities and limitations .....</b>	<b>83</b>
1. Introduction.....	83
2. Rational choice and behavioural economics: theory and policy implications .....	84
2.1. The traditional approach: rational choice theory .....	84
2.2. Behavioural economics: new insights and new policy implications.....	87

3.	The ties between behavioural economics and rational choice and their implications .....	91
4.	Objections to behavioural economics and its use in policymaking: the critic from the fast and frugal heuristics approach.....	93
4.1.	Boost policies: characteristics and justification .....	97
4.2.	Is there a real dispute? Core claims vs rhetorical claims .....	98
5.	Concluding remarks.....	101
<b>Chapter 4 A pragmatic definition of nudges.....</b>		<b>103</b>
1.	Introduction.....	103
2.	Problems with some of the original properties .....	104
3.	A pragmatic definition of nudges .....	106
4.	Final remarks .....	106
<b>Chapter 5 Attitudes towards nudges: a literature review.....</b>		<b>108</b>
1.	Introduction.....	108
2.	Attitudes towards nudges: an overview.....	109
2.1.	Overview of the methods and characteristics.....	110
2.2.	Overview of the findings.....	117
3.	Unresolved questions and future directions.....	129
3.1.	Elements of the design and implementation and nudges acceptability.....	131
3.2.	Framing and nudge acceptability .....	134
4.	Concluding remarks.....	136
<b>Chapter 6 Public acceptability of nudges in Spain: a nationally representative survey.....</b>		<b>137</b>
1.	Introduction.....	137
2.	Research questions and hypothesis.....	138
3.	Materials and methods.....	141
4.	Results.....	145
4.1.	Do people like nudges in Spain?.....	145
4.2.	Which nudges do people support? .....	149
4.3.	Which factors predict the acceptability of nudges? .....	154
5.	Discussion.....	160
6.	Concluding remarks.....	165

<b>Chapter 7 Preferences for automatic and reflective nudges: experimental evidence and implications .....</b>	<b>166</b>
1. Introduction.....	166
2. Research questions and hypothesis.....	167
3. Study A. Are preferences for automatic nudges affected by how nudges are described?.....	172
3.1. Materials and Methods.....	172
3.2. Results.....	176
4. Study B. How does the context of choice affect the acceptability of automatic nudges? .....	186
4.1. Materials and Methods.....	186
4.2. Results.....	190
5. Discussion.....	195
6. Concluding remarks.....	204
<b>Conclusions.....</b>	<b>205</b>
1. Main contributions.....	205
2. Limitations and possibilities for future research .....	209
3. Final thought.....	210
<b>References.....</b>	<b>211</b>
<b>Appendix A.....</b>	<b>221</b>
1. Survey and questionnaire.....	221
2. Data preparation for the survey .....	230
3. Co-variables descriptive statics .....	233
4. Data on perceived effectiveness and perceived manipulation .....	236
<b>Appendix B.....</b>	<b>238</b>
1. Supplementary materials Study A. ....	238
2. Supplementary materials Study B. ....	257

## List of tables

Table 1. Relationship between the arguments and properties of nudges.....	15
Table 2. Counterarguments against and problems related to the original nudge characterisation. ....	21
Table 3. Summary of classifications of nudges and their ethical implications.....	51
Table 4. Types of classifications of nudges according to their type of criteria. ....	54
Table 5. Features of nudges interventions. ....	73
Table 6. Domains and types of interventions considered in the attitudes towards nudges literature. .	113
Table 7. List of 17 nudges, categorised by type of nudge and aim of nudge .....	142
Table 8. Sample sociodemographic characteristics .....	144
Table 9. Logistic regression models predicting no-acceptability of the 17 nudges.....	155
Table 10. Interventions included in the survey of Study A .....	173
Table 11. Summary of treatments and experimental conditions of Study A.....	174
Table 12. Logistic regression models to assess the impact of conditions on acceptability. ....	180
Table 13. Logistic regression models predicting the acceptability for the eight interventions. ....	183
Table 14. Summary of treatments and the eight experimental conditions of Study B. ....	188
Table 15. Survey and questionnaire for survey .....	221
Table 16. Sample characteristics of participants in the survey .....	233
Table 17. Experimental sequence and questions Study A.....	238
Table 18. Sample characteristics participants Study A.....	250
Table 19. Contrastive vignettes used in the healthy eating scenario (A).....	257
Table 20. Contrastive vignettes used in the organ donation scenario (B) .....	258
Table 21. Experimental sequence and questions Study B .....	259
Table 22. Sample characteristics participants Study B.....	262

## List of figures

Figure 1. Acceptability of the 17 nudges by response.....	145
Figure 2. Pro-self vs Pro-Social nudges .....	150
Figure 3. Automatic vs Reflective nudges.....	151
Figure 4. Acceptability of the eight interventions by response. ....	177
Figure 5. Means for each response variable by condition in the healthy eating scenario. ....	192
Figure 6. Means for each response variable by condition in the organ donation scenario.....	193
Figure 7. Perceived effectiveness for the 17 nudges. ....	236
Figure 8. Perceived manipulation for the 17 nudges. ....	237
Figure 9. Perceived effectiveness of the eight interventions by response. ....	255
Figure 10. Perceived manipulation of the eight interventions by response.....	256

# Introduction

Nudges are interventions in the decision-making environment that steer individuals in a particular direction without altering their economic incentives or ruling out any options. Their implementation is based on evidence that shows that decision-making is systematically affected by heuristics and cognitive biases; agents have non-consistent preferences that vary depending on the context of choice, and behaviour is very sensitive to social influence. Bearing these factors in mind, nudges work by making some options more attractive or easier in cognitive and social terms but still allowing people to choose what they want (Thaler & Sunstein, 2008). Thaler and Sunstein first introduced the concept in their book *'Nudge, improving decision about health, wealth and happiness'*. In it, the authors discuss how to use nudges in policymaking and base their proposal on three main pillars: behavioural economics, *choice architecture* and *libertarian paternalism*.

Thaler and Sunstein proposal is based on inputs from behavioural sciences, behavioural economics and cognitive psychology about how people make decisions in real contexts. Their approach relies on empirical evidence that rejects the standard assumptions of rational behaviour in neoclassic economics and recognises that people are not rational beings, and do not always behave like *homo economicus* or *Econs*. Empirical evidence in the social sciences indicates that agents' decisions are systematically affected by cognitive biases and social emotions, and strongly influenced by the context of choice. The evidence suggests that people behave like *Humans* as a result of these behavioural biases, i.e., they struggle to make good choices and to behave according to their intentions, which harms their wellbeing.

In *'Nudge'*, Thaler and Sunstein acknowledge and discuss the challenges and opportunities that these findings pose for public policy. The authors claim that policymakers ought to interfere with people's choices to prevent negative consequences and make people better off, building on evidence that shows that people often make bad decisions for themselves as a result of these factors. At the same time, they introduce the idea that these findings are useful for policymaking. They contend that, by using behavioural insights, policymakers can improve the design and performance of policy interventions and employ a new tool to drive behavioural change: nudges.

In addition to the practical value of nudges, Thaler and Sunstein highlight two aspects that make them a legitimate tool for policy intervention. Thaler and Sunstein embed the use of nudges within libertarian paternalism, a normative framework that wants to influence behaviour to promote individual wellbeing while respecting freedom of choice. The authors maintain that people's decisions suffer due to the effect of behavioural biases. Following this evidence, they argue that nudges promote people wellbeing 'as judged by themselves' (Thaler & Sunstein, 2008, p. 5) by steering people towards choices that they would have made if it were not for the effect of cognitive limitations.

At the same time, Thaler and Sunstein argue that in general, people should be free to do what they like. Consequently, they contend that nudges are the best option for interfering in people's choices, because they allow people to go their own way. According to the authors, nudges have a choice preserving nature; unlike traditional instruments to influence behaviours, such as taxes and regulation, nudges aim to change behaviour without implying coercion. Instead of altering economic incentives or ruling out some of the options in a particular context of choice, nudges only change the presentation of these options, making some more or less likely, but still allowing people to choose whatever they want. In that respect, nudges are deemed to be ethically appropriate because they promote those options that are best for people, while still allowing those who want to go their own way to do so.

Using a combination of theoretical and normative arguments, Thaler and Sunstein show that, by understanding behaviour, policymakers can promote effective behavioural change and design choice environments that steer people towards beneficial directions without restricting their freedom of choice. These elements give nudges an unusual appeal for policymaking. Their empirically grounded nature means that their implementation appears to be a useful strategy for influencing behaviour in different domains. Furthermore, the fact that they respect freedom of choice and promote individual and social wellbeing brings nudges acceptability from all sides of the political spectrum.

This idea quickly gained the interest of policy practitioners after the publication of '*Nudge*'. In 2010, the United Kingdom set up The Behavioural Insight Team (BIT), unofficially 'The Nudge Unit', to practice and promote the use of behavioural insights and nudges in public policy. In the US, Cass Sunstein became head of the Office of Information and Regulatory Affairs from 2008 to 2012 in the first Obama administration. In 2015, they created the Social and Behavioural Sciences Team (SBST). Following the example of the BIT, the SBST also worked on applying lessons from behavioural sciences into policymaking. The interest in nudges and behavioural insight in policymaking took off after these two pioneering experiments.

In the last decade, nudge implementation has gained institutional recognition and given a significant boost to the establishment of *behaviourally informed policy*. The approach has attracted the interest and approval of governments of different ideologies and international public policy organisations. Nowadays, more than 200 organisations have adopted the use of nudges (OECD, 2017). Public administrations around the world use behavioural insights and implement nudges. Likewise, organisations such as the OECD, the World Bank, and the European Commission embrace behavioural insights and have teams working on the topic (Afif et al., 2018; Lourenço et al., 2016; World Bank, 2015). Following this trend, a growing number of independent institutions and "nudge units" specialise in implementing policy changes through nudges. For instance, since opening in 2010, the BIT has grown significantly and now works as an independent company, with offices in London, New York, Shanghai and Toronto (BIT, 2019). The interest in nudges is also palpable within academia and applied research

institutions, with an increasing number of papers, projects and courses on behavioural insights and nudges

Ten years after their debut, the evidence on the performance of nudges indicates that they are useful when it comes to tackling policy problems in several domains. Some examples of nudges with successful results include the implementation of default rules, the communication of social norms, the framing of information, the simplification of information, the re-arranging of the physical environment, the use of reminders and warnings, the provision of feedback and the use of self-regulating strategies. In general, the evidence on the performance of nudges indicates that they produce aggregate impact at a low cost and are useful when it comes to tackling policy challenges of different natures. As a result, the implementation of nudges is seen as a valuable policy tool to promote effective and ethical behavioural change.

Nevertheless, despite their growing implementation, there is also considerable debate about whether nudges are an appropriate policy instrument. The use of nudges raises objections and their consolidation as a policy instrument faces legitimisation and conceptualisation issues.

The first striking critique is that nudges are unethical. The acceptability of nudges as a policy tool is questioned by several critics, who oppose their normative acceptability. For instance, the implementation of nudges faces objections of overreaching and illegitimate paternalism, undermining people's wellbeing and amplifying the nanny state. Likewise, nudges face claims of being manipulative and non-transparent, able to steer individuals towards a choice without their awareness or consent and are often described as a potential threat to agents' freedom of choice, autonomy and dignity due to the way in which they change behaviour. These arguments point towards nudges' limitations in terms of respecting freedom of choice and raise worries about their use being a coercive strategy. Likewise, nudges face accusations of covert paternalism, and being a tool that could easily be used for evil purposes. While the topic has been widely discussed in literature, the ethical acceptability of nudges is still very much an unresolved issue, and more discussion and evidence are, therefore, needed to clarify their moral profile.

The second important critique about nudges is directed towards the underlying theoretical arguments and their implications in policymaking. Some authors of research into judgment and decision-making suggest that the implementation of nudges relies on a narrow and negative understanding of human rationality that lacks validity when it comes to understanding decision-making and it is not useful to design strategies for behavioural change. Some empirical evidence supports these objections, since nudges show inconsistent results about their effectiveness. For instance, similar interventions perform differently within different social and cultural contexts, and some interventions are ineffective depending on the domain in which they are applied. Likewise, some authors also worry about nudges failing to have persistent effects, being useless in the long-term and having unintended negative

consequences. These objections question the practical value of nudges in policymaking. Overall, their acceptability as a policy tool is challenged by competing interpretations of rationality and different understandings of how behavioural insights should inform policymaking.

A final challenge to the consolidation of nudges as policy instruments is the lack of agreement on *what nudges are* and *what nudges involve*. Much of the discussion about the problems posed by nudges and their implications arise from confusion about their meaning. Despite the widespread use of the concept *nudge*, there is not a unique understanding of it. The concept does not have a substantive definition, and authors tend to define it by giving examples. In literature, the term *nudge* is commonly used as a generic label to discuss interventions of different types and contexts of applicability, often straying into debates about behaviour, rationality and public policy. It is challenging to articulate which interventions are or are not nudges because their essential characteristics as a policy intervention are not precise. As a result, while literature contains extensive discussions on nudges and their acceptability, these contributions often employ different and vague understandings of the term *nudge*.

In these circumstances, the acceptability of nudges is in doubt. The implementation of nudges faces several challenges: the different objections trigger reservations about their ethical implications and uncertainties about their practical value as a policy tool. Both issues are aggravated by the lack of consensus over the meaning of the concept *nudge*, making the discussion of these objections even more challenging. At the same time, positive experiences with nudges indicate that they have value as policy instruments when used to bring about behavioural change and tackle policy challenges.

In light of the overall debate around nudges, the question of whether *to nudge or not to nudge* is unresolved. There are arguments for and against nudges, and the debate is inconclusive. Proponents of the approach defend nudges, stressing their empirical base and their commitment to libertarian paternalism. Critics argue that these arguments are insufficient to overcome their objections and oppose the use of nudges to modify citizens' behaviour. In this context, generally arguing in favour or against nudges is almost pointless. Instead, it appears more important to explore what this discussion says about *how* to nudge in policymaking.

The debate around nudges indicates that nudges have both potential and limits. They can be useful, but the controversy around their use requires specific attention to counter any ethical doubts and doubts about their effectiveness, and attention also needs to be paid to issues of conceptualisation. Thaler and Sunstein's proposal is attractive; however, it has several problems related to its intended performance. This thesis explores the main objections to using nudges in policymaking and discusses crucial issues about their acceptability as a policy tool. It argues that nudges have value for public policy, but the concept needs to be revised.

## **1. Research objectives and questions**

This thesis explores the acceptability of nudges by discussing the advantages, limitations and challenges of using nudges in policymaking. Two research objectives guide the research: the first is addressing the main problems and objections related to nudges; the second is to provide a conceptualisation of nudges that focuses on their practical value as a policy instrument for behavioural change. The thesis has two main parts: firstly, a theoretical part, which consists of a conceptualisation of nudges as a policy tool; secondly, an empirical part, which includes an analysis of the public acceptability of nudges.

The discussion in both sections is intended to respond to four main questions:

- How should nudges be understood as valuable policy instruments?
- Are nudges ethically acceptable? Why?
- Are the theoretical bases of nudges useful for public policy?
- What are the main challenges related and limitations to the acceptability of the use of nudges? How can they be resolved?

The first part of the thesis discusses the characterisation of nudges as a policy instrument. This part includes a discussion on the concept of a nudge, focusing on conceptual, ethical and theoretical issues. This thesis looks at the original notion of nudges, their theoretical and empirical bases and the main criticisms in both dimensions. The discussion goes on to outline problems with Thaler and Sunstein's notion of nudges and proposes an alternative understanding of nudges, which advocates dropping the normative connotations of nudges and focuses on their practical value as a policy tool.

The second part includes three empirical studies on attitudes toward nudges: a survey of the public acceptability of nudges in Spain, and two experimental studies that also consider issues related to their acceptability. The empirical part addresses some of the concerns discussed in the theoretical debate and emphasises that the discussion on the acceptability of nudges as policy tools must include considerations about whether people like nudges, which nudges they prefer and in which context they accept their use. To answer these questions, the thesis looks at arguments in favour and against nudges and offers some ideas and conclusions on how nudges could be useful in policymaking.

## **2. Outline of the thesis**

The thesis is structured in eight chapters:

Chapter 1 presents Thaler and Sunstein's original nudge characterisation and details the main problems with and lines of criticism of their approach. The aim of the chapter is twofold: it outlines the origins of nudges and the main problems with them. The first section explores the main ideas and arguments in '*Nudge*' and outlines the explicit and implicit properties that the authors assign to nudges within the original characterisation. The second section looks at the main problems and objections of this approach as a way of introducing the themes that will be addressed in later chapters.

Chapter 2 explores the ethical acceptability of nudges. The chapter addresses objections to the aims of and justifications for nudges (aims of nudges) and objections to how nudges influence choices (nudging mechanisms). The chapter argues that the discussion on the ethical acceptability of nudges overstates the negative implications of nudges and has tended to overlook elements that play in favour of their normative legitimisation as a policy instrument. Accordingly, it argues that nudges are not intrinsically morally problematic; instead, their moral acceptability as a policy instrument depends on three aspects: the nudge itself, the context of use, and the aim that it promotes. The chapter discusses the ways in which these three elements shape the acceptability of nudges.

Chapter 3 discusses the use of behavioural insight in policymaking. The implementation of nudges is associated with behavioural economics and carries a specific conception of rationality and decision-making. Alternative frameworks on research into judgment and decision-making question findings from behavioural economics and, therefore, also question the use of nudges in policymaking. Chapter 3 addresses these objections by exploring the bases and problems of behavioural economics and looking at how behavioural economics and other theories about behaviour and decision-making inform the design and implementation of policy interventions. The chapter contends that insights from behavioural economics are useful for policymaking but should not be used in an uncritical way.

Chapter 4 presents a new idea about nudges. The chapter integrates the themes and arguments discussed in chapters 1, 2 and 3 and argues that the concept of nudges should not include normative connotations. Accordingly, it defines nudges as a policy instrument with specific characteristics and practical value for behavioural change but without pre-defined ideas about the values and aims that it should promote. It then discusses some of the potential negative implications of nudges and argues the idea that their role as a public policy tool should not only be informed by theory but also by people's opinion on nudges.

Chapter 5 presents research on public attitudes towards nudges. It reviews the available literature on the topic, summarises the main findings, and identifies gaps and questions that are still unresolved.

Chapter 6 analyses public opinion towards nudges in Spain. Following the discussion in Chapter 5, it presents the results of a nationally representative survey ( $n= 617$ ,  $e=3.88$ ) that analyses how people respond to the use of different nudges to tackle different policy problems. It also compares these results with the findings from other countries. The results indicate that, in general, Spaniards support the use of nudges, but the acceptance varies depending on the objective and the type of nudge.

Chapter 7 presents two online experiments that focus on assessing preferences for two different types of nudges: *automatic nudges*, which are less noticeable and work with no deliberation and awareness, and *reflective nudges*, which are noticeable, work by engaging deliberation and produce a more reflective response. The first study is a survey-experiment that explores whether preferences for automatic and reflective nudges vary when subjects are informed about the effectiveness of the nudge

and the nudge's potential to work without behavioural awareness. The results indicate that this information does not change people's preference for nudge interventions. The second study is a vignette experiment that explores whether the acceptability of nudges varies when subjects know about the effect that the context of choice has on choices and whether this varies in different domains. The results indicate that people do not oppose the use of automatic nudges, but their acceptability depends on the aim they are promoting and the domain in which they are applied. The chapter advances the discussion on nudge acceptability and offers insights on folk conceptions of choice, decision-making and autonomy. The findings are relevant when it comes to deciding how to nudge, where to apply nudges and how to communicate their use.

The thesis ends with the conclusions.

# Chapter 1

## The idea of nudge: origins, foundations and problems

### 1. Introduction

The term nudge has an ambiguous meaning. In *'Nudge'* Thaler and Sunstein (2008) define nudges as interventions in the choice architecture that steer individuals towards a particular direction without altering their economic incentives and without forbidding any options. Using this definition, the authors offer different examples and discuss how nudges can be implemented to achieve positive outcomes in many domains. There are, however, two main problems with their understanding of nudges. Firstly, the authors do not offer a substantive definition of the term. The lack of substantive criteria has created confusion over which interventions are or are not nudges. As a result, different interventions, with different characteristics and use, may be categorised as nudges or not depending on who classifies them and how they understand or adopt the term. Secondly, while Thaler and Sunstein present nudges as a general instrument to influence people's choices, their understanding of nudges relies on specific theoretical and normative arguments. The authors associate the use of nudges with libertarian paternalism, a normative framework that aims to improve individual wellbeing and respect freedom of choice. Likewise, their approach relies on behavioural economics and carries a specific understanding of human rationality. Both libertarian paternalism and behavioural economics characterise the original understanding of the concept of nudges.

Throughout this thesis, the term *original nudge characterisation* will be used to refer to Thaler and Sunstein's understanding of nudges. I argue that the original nudge characterisation defines nudges as a unitary concept with five properties. A nudge:

- i. does not change economic incentives
- ii. does not forbid any options
- iii. affects the behaviour of Humans and is ignored by Econs
- iv. improves people's wellbeing as judged by themselves and
- v. respects freedom of choice.

The properties define nudges as a policy instrument that has both a practical and normative value for pursuing behavioural change.

Even though Thaler and Sunstein's approach is based on empirical evidence, the original nudge characterisation and the arguments that support it have been subjected to considerable criticism. Several authors identify problems with the arguments underlying the approach and stress that nudges often struggle to comply with all their properties. At the same time, these objections create conceptualisation problems. As a result, the term nudge is often used imprecisely and beyond its original meaning. It

references different interventions (some of which include the five properties and some that do not) and is employed to introduce debates on rationality, ethics and public policy. In general, the term is employed uncritically without attention to what is being referenced when using the concept nudge. At the same time, the discussion on nudges is profoundly shaped by elements of its original characterisation. Thereby, while the answer to what exactly a nudge is remains diffuse and controversial, it always starts or considers features and problems with Thaler and Sunstein's original notion.

Chapter 1 revises the original nudge characterisation, discusses its main arguments, and outlines its main problems. The chapter serves to introduce the topics that further chapters will discuss in depth. It aims to understand the starting points for nudges and their problems, while subsequent chapters will look at how they can go on to be used as a policy tool. The chapter proceeds as follows: Section 2 outlines the explicit and implicit properties of nudges. It addresses the association between nudges and behavioural economics and then the association between nudges and libertarian paternalism. Section 3 outlines the main problems with the arguments supporting the original nudge characterisation and in which ways these objections challenge the definition of nudges and their value as a policy tool. The chapter ends with section 4, which remarks that the original nudge characterisation has unresolvable issues and emphasises the need for a new understanding of the concept of nudges that drops the normative properties.

## **2. The original nudge characterisation**

Thaler and Sunstein's original definition posits that nudges are '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. Nudges are not mandates. Putting fruit at eye level counts as a nudge. Banning junk food does not'. This definition serves to distinguish nudges from traditional policy instruments; as opposed to taxes and bans, nudges do not change economic incentives and maintain all the options available in a particular context of choice. The definition provided by Thaler and Sunstein outlines two definitional properties of nudges interventions:

- i. Nudges do not significantly alter economic incentives
- ii. Nudges do not forbid any option available in a particular context of choice.

Thaler and Sunstein use this definition as the base to present nudges and to discuss their uses and implications. However, they associate the use of nudges with a particular understanding of government interference: libertarian paternalism. Likewise, their arguments build upon a specific notion of human rationality based on behavioural economics. Both frameworks, libertarian paternalism and behavioural economics, serve the authors to develop and defend their approach to behavioural change, with

particular ideas on which ends, and values nudges should promote. The following sections revise in which ways libertarian paternalism and behavioural economics characterise the notion of a nudge.

## **2.1. Nudge and behavioural economics**

The first key idea in *'Nudge'* is the notion that people have recurrent problems in making good decisions for themselves. Thaler and Sunstein explore evidence from behavioural sciences, especially from behavioural economics, that shows that individuals are not perfectly rational agents, and therefore are not always prepared to make decisions that maximise their wellbeing. The authors describe substantial evidence on real-life behaviour that challenges the descriptive power of rational choice theory and shows that individuals rarely behave like Econs; this is perfectly rational, with stable preferences, unlimited computation capabilities and perfect self-control. Instead, individuals behave like Humans and exhibit systematic and predictable decision-making flaws.

Thaler and Sunstein further characterise their understanding of rationality with Kahneman (2011) account of the dual-process cognitive theory. Kahneman (2011) describes a human cognition that works by employing two distinct cognitive systems: System 1 and System 2. System 1, or The Automatic System, is fast, uncontrolled, unconscious, intuitive but affected by heuristics and emotions. System 2, or The Reflective System, is deliberative, conscious, rational, controlled, but slower (Kahneman 2011). System 1 is more likely to make mistakes; System 2 is usually right but not always available. Because of the effect of System 1, people often fail to behave like Econs. Psychological mechanisms associated with System 1 interfere with their decision-making process, making people behave like Humans and steering them towards irrational decisions that burden their wellbeing.

In *'Nudge'*, Thaler and Sunstein summarise different psychological mechanisms that can have negative effects on people's choices in three groups: biases and blunders, temptation, and following the herd.

The first group, *biases and blunders*, describes common rules of thumb in decision making which, while useful for producing rapid decisions, often lead to incorrect judgments. Following experimental findings from the heuristics-and-biases research programme (Tversky & Kahneman, 1974), Thaler and Sunstein describe people's difficulties in making good judgments of probability under situations of risk and uncertainty. Heuristics such as anchoring, availability, and representativeness, and associated biases such as overconfidence and optimism bias affect the decision-making process. Likewise, prospect theory research (Kahneman & Tversky, 1979) indicates that people are highly sensitive to framing effects. Their decisions differ greatly depending on the description of options, especially if these options highlight losses rather than gains; they are generally loss-averse and have the tendency to favour the status quo. These factors showcase that individuals do not weigh options following the expected utility model, and often settle for non-optimising choices.

The second group, *temptation*, illustrates people's difficulties in undertaking plans, controlling impulses, and deciding between options with conflicting immediate and future rewards. Research in this regard suggests that people make plans with the intention of sticking to them yet, in many cases, they fail to follow through. They struggle due to the *hot-cold empathy gap*: in a hot state, they are unable to resist temptation, while in the cold state, they underestimate the effect that temptation will have on their behaviour. Under the 'temptation' category, Thaler and Sunstein mention extensive literature on problems of self-control, and context-dependent preferences, both of which pose important challenges to the rational choice theory assumption of consistent preferences.

The third group, *following the herd*, focuses on the role that social influence and social norms have in everyday decision-making. The authors refer to extensive research that shows that people usually conform to what others do. Research on social influence shows that people's behaviour is affected by social norms, social conventions and social preferences and does not reflect the assumptions of self-interest and egoism present within the rational choice theory framework. The behaviour of others gives information about what is the correct thing to do, and how to behave to avoid disapproval and this information has a significant effect on people's behaviour. However, in some cases, following the herd results in situations that undermine individual self-interest and can potentially lead people towards suboptimal choices.

With the examples of these three groups, Thaler and Sunstein give theoretical consistency to the idea that real decision-makers, or Humans in their terms, exhibit limited self-interest, limited will power, preference inconsistency due to time and situation, and difficulties in processing all the available information due to cognitive constraints. Likewise, they emphasise that the effect of behavioural biases is systematic, predictable and generally unavoidable. In that respect, throughout '*Nudge*', the authors provide evidence that real-life behaviour differs significantly from rational choice theory expectations.

The second key idea in '*Nudge*' is the explicit discussion on how to incorporate insights about human behaviour within public policy. In light of behavioural economics findings, Thaler and Sunstein stress the need to revise how we understand behaviour and how we approach behavioural change. According to Thaler and Sunstein, the evidence that documented and systematic behavioural biases affect, and compromise people's decisions is of great interest to improving public policy. The authors argue that the evidence from behavioural economics presents new considerations regarding when it is necessary for governments to intervene to improve people's lives. In '*Nudge*', they stress that behavioural biases systematically compromise people's wellbeing and prevent them from living a happy, healthy and wealthy life. Accordingly, they argue that choice architects, i.e., governments and institutions with the power and responsibility to organise the contexts in which people make decisions, should interfere with these decisions to prevent poor choices and promote those that improve people's lives. As they explicitly state: 'If you indirectly influence the choices other people make, you are a choice architect. And since the choices you are influencing are going to be made by Humans, you will want your

architecture to reflect a good understanding of how humans behave. In particular, you will want to ensure that the Automatic System doesn't get all confused' (Thaler & Sunstein, 2008, p. 83).

While their approach defines a new challenge for public policy, it also offers a novel approach to solving it. Thaler and Sunstein argue that behavioural insights are useful when thinking about new tools for behavioural change. Thaler and Sunstein suggest that choice architects should take into consideration that people's behaviour is highly influenced by heuristics and biases, social influence, and the context of choice. Thereby it is possible to leverage these factors and use nudges to motivate behavioural change.

Using behavioural economics as a base, Thaler and Sunstein provide another definitional property of nudges:

- iii. Nudges are 'any factor that significantly alters the behaviour of Humans, although it would be ignored by Econs' (Thaler & Sunstein, 2008, p. 8).

Property (iii) establishes that to nudge means to promote behavioural change by preventing, targeting or benefitting from behavioural biases. Following property (iii), the authors consider the implementation of default options, the communication of social norms, the provision of feedback, and the framing of information as nudges that can be implemented to pursue behavioural change. These interventions are related to psychological mechanisms of biases and blunders, temptations and following the herd, and as such, can be classified as nudges.

## **2.2. Nudge and libertarian paternalism**

The third central idea in '*Nudge*' relates to its value and goals as a behavioural change strategy. Thaler and Sunstein associate the use of nudges with libertarian paternalism, an approach, which they argue, 'can preserve freedom of choice while also nudging people in directions that will improve their lives' (Thaler & Sunstein, 2008, p. 252). Libertarian paternalism establishes two criteria to shape individuals' choices. The paternalistic side of the approach aims to change individuals' behaviour to 'make choosers better off, as judged by themselves' (Thaler & Sunstein, 2008, p. 5). The libertarian side of the approach, however, wants to do so without restricting the original set of choices, thereby respecting individuals' freedom of choice.

The paternalistic aspect of the approach stems from its intention to nudge individuals towards decisions that will make them better off. Building on behavioural economics, Thaler and Sunstein argue that people often engage in non-optimising, non-rational choices with negative effects for their wellbeing. To prevent these choices, they say that choice architects should use behavioural insights to nudge people in directions that will improve their lives. However, their justification differs from traditional arguments in favour of paternalistic interventions. Thaler and Sunstein argue that nudges promote individuals' wellbeing 'as judged by themselves' (Thaler & Sunstein, 2008, p. 5), i.e. in directions agents agree with

and would have chosen if they were not affected by cognitive biases, bounded self-control, and limited time, information, and cognitive abilities. In that respect, they defend that nudges are legitimate since they steer people towards those options that their rational self would have chosen free from the influence of irrational and undesired biases.

Reinforcing the idea that choices do not always reflect individuals' rational preferences, Thaler and Sunstein further defend the use of nudges under the argument that choice architecture is inevitable. The authors argue that when choice architects organise the context in which people make decisions, they cannot avoid influencing these choices. They emphasise that there is no neutral design, options are always arranged in one way or another, and the way they are arranged will always affect behaviour. Since it is not possible to avoid influencing people's choices, Thaler and Sunstein argue that rather than allowing arbitrary distributions to shape decisions towards adverse outcomes, it is better to organise the context of choice in ways that nudge people towards good choices. In their view, paternalistic intervention with individual choices is necessary to prevent rationality failures. Likewise, this interference is generally unproblematic because people agree with the direction in which they are being nudged.

Despite defending the need to interfere with people's choices for their own good, Thaler and Sunstein also defend that nudges ought to respect freedom of choice. The libertarian aspect of the approach stresses that nudges should always be liberty-preserving and 'make it easy for people to go their way' (Thaler & Sunstein, 2008, p. 5). The authors emphasise that they approach 'attempts to design policies that increase or maintain freedom of choice'. The key idea is that nudges are designed to improve the likelihood of specific options, but do not block or forbid any options or significantly dismiss the possibility to opt-out of the nudge. Since nudges do not restrict the number of choices and do not change economic incentives, individuals are free to choose any other option besides the one intended by the nudge; thus, they preserve freedom of choice. With this argument, the authors justify why libertarian paternalism does not imply coercion and reconcile paternalistic intention with the value of allowing people to do their own thing if they want to.

By combining paternalistic and libertarian values, Thaler and Sunstein (2003, 2008) provided a normative framework that promotes soft interventions—in the form of nudges—in those domains in which both individual and social wellbeing is threatened due to behavioural biases. Accordingly, they argue that affecting choices through nudges is generally unproblematic because nudges steer people in directions they agree with and do so by maintaining all the options and preserving individuals' ability to opt-out of the nudge.

The association of nudges with libertarian paternalism assigns two more definitional properties to nudges:

- iv. Nudges tackle rationality failures and promote individual well-being as judged by the individuals themselves.
- v. Nudges respect freedom of choice.

### **2.3. The original nudge characterisation, five definitional properties.**

Reviewing the main ideas in *'Nudge'* unveils the original characterisation of nudges. With the primary definition, Thaler and Sunstein define nudges with two definitional properties: (i) nudges do not significantly change economic incentives and (ii) nudges do not forbid any options. Based on the theoretical framework underlying the approach, they add a third property (iii) nudges affect the behaviour of Humans and would be ignored by Econs. Relying on these aspects, Thaler and Sunstein characterise the idea of nudges as policy instruments. Nudges are distinct from traditional instruments for changing people's behaviour because they do not alter incentives or opportunities (definitional properties (i) and (ii)). Instead, they work by exploiting or counteracting specific psychological mechanisms (definitional properties (iii)). These definitional properties also allow Thaler and Sunstein to defend the ethical value of nudges. As a result of definitional properties (i) and (ii), the authors maintain that: (iv) nudges do not imply coercion and respect people's freedom of choice; as a result of definitional property (iii), the authors maintain that (v) nudges correct irrational decisions and have a benevolent intention: to make people better off and to promote their true rational preferences.

Following the arguments within the original nudge characterisation, *a nudge* is defined as a behavioural change intervention with five properties:

- i. Nudges do not significantly change economic incentives
- ii. Nudges do not forbid any options
- iii. Nudges affect the behaviour of Humans and are ignored by Econs
- iv. Nudges tackle rationality failures and promote individual well-being as judged by the individuals themselves.
- v. Nudges respect freedom of choice.

A nudge is defined as a unitary concept and as a policy instrument that has practical and normative value to achieve behavioural change.

### **3. Problems with the arguments that support the original nudge characterisation**

To support the original nudge characterisation, Thaler and Sunstein rely on specific arguments on rationality and decision-making and influence and interference. The main arguments underlying their approach can be summarised as follow:

- (A1) People's behaviour is frequently affected by behavioural biases.
- (A2) Biases are systematic, predictable and generally unavoidable.

- (A3) Biases produce rationality failures and lead people towards choices that undermine their wellbeing.
- (A4) People have rational preferences.
- (A5) Reducing the number of choices and not altering economic incentives preserves freedom of choice.

Embracing these arguments is required in order to accept Thaler and Sunstein’s understanding of nudges, and it is essential for defining nudges as a tool with the properties described above. Arguments 1 and 2 are key to supporting property (iii). The fact that behaviour is systematically affected by behavioural biases means that nudges can be used to change behaviour, and they will affect the behaviour of Humans rather than the behaviour of Econs. Arguments 3 and 4 support property (iv). The fact that biases produce rationality failures and the idea that people have rational latent preferences are essential to the argument that nudges correct rationality failures and promote individual well-being as judged by the individuals themselves. Finally, Argument 5 is critical to supporting property (v); the understanding that, by not reducing the options and not changing economic incentives, people still can choose freely, is vital to establishing nudges as interventions that respect freedom of choice. Table 1 summarises the arguments that support the properties of nudges.

**Table 1. Relationship between the arguments and properties of nudges.**

<b>Arguments</b>	<b>Properties of nudges</b>
(A1) People’s behaviour is frequently affected by behavioural biases. (A2) Biases are systematic, predictable and generally unavoidable.	(iii) Nudges affect the behaviour of Humans and are ignored by Econs.
(A3) Biases produce <i>rationality failures</i> and lead people towards choices that undermine their wellbeing. (A4) People have <i>rational preferences</i>	(iv) Nudges tackle rationality failures and promote subjective individual wellbeing.
(A5) Not reducing the number of choices and not altering economic incentives preserves freedom of choice	(v) Nudges preserve freedom of choice

Despite the consistency of the approach and the evidence that underpins each argument, the arguments within the original nudge characterisation have been subjected to considerable criticism and, some of the objections pose direct challenges to the properties of nudges. Notoriously, problems with the underlying understanding of human rationality and the conception of freedom of choice within the approach raise doubts about whether nudges can comply with properties (iii), (iv) and (v). The sections below outline the main objections to the core arguments and describe how they challenge the original

nudge characterisation. First, I concentrate on objections regarding the understanding of rationality, and then I tackle objections related to the influence of nudges and freedom of choice.

### **3.1. Problems with the conception of rationality**

Thaler and Sunstein perceive nudges as interventions that (iii) affect the behaviour of Humans and are ignored by Econs and (iv) that tackle rationality failures and improve wellbeing. To argue that nudges comply with these properties, Thaler and Sunstein rely on the following arguments:

- (A1) People's behaviour is frequently affected by behavioural biases.
- (A2) Biases are systematic, predictable and generally unavoidable.
- (A3) Biases lead people towards choices that undermine their wellbeing.
- (A4) People have rational preferences and would like to avoid the effect of behavioural biases

The existence of behavioural biases, the confirmation that their effect is systematic, predictable and unavoidable, and the idea that they compromise people's choices and welfare and do so against their own will, are core arguments to sustaining the original nudge characterisation. Relying on these arguments Thaler and Sunstein maintain that nudges can effectively change behaviour and, in doing so, promote people's best interest. Since people's behaviour is affected by behavioural biases (A1) and these biases are systematic predictable and unavoidable (A2), it is possible to detect when and in which context they take place and re-structure the context of choice to neutralise or correct their effect (property iii). Likewise, the fact that behavioural biases cause rationality failures and undermine people wellbeing (A3) serves as an argument in favour of interfering with people's choices. Finally, by supporting the premise that people want to make rational decisions but struggle to do so due to behavioural biases (A4), Thaler and Sunstein claim that nudges promote people's best interest as judged by themselves (property vi). Argument 1 and Argument 2 support the pragmatic value of nudges, Argument 3 and Argument 4, its legitimisation as a behavioural change strategy.

To sustain arguments 1 to 4, Thaler and Sunstein rely on behavioural economics. As already commented, the evidence from behavioural economics shows that individuals systematically deviate from the standard model of rationality. Decision-making is regularly affected by behavioural biases; people have non-consistent preferences and cognitive limitations impede their reasoning. Behavioural economics also maintains that these limitations compromise people's wellbeing and cause *internalities*, preventing people from making those decisions that they would like to make. However, problems arise from the theoretical behavioural economics framework, and several authors suggest that there are empirical and theoretical limitations to sustaining arguments 1, 2, 3 and 4.

Many authors identified problems with understanding the rationality underlying the original nudge characterisation. A priori, behavioural economics presents itself as a challenge to neoclassical economics. It recognises its limited descriptive power and advocates for improving the explanatory

power of the theory by including more realistic and plausible social and cognitive foundations. However, rational choice theory still has a defining role within behavioural economics. Behavioural economics uses neoclassical theory as a benchmark, as the normative background that outlines how rational behaviour *ought to be*. Thereby, the neoclassical behaviour model has a substantive effect on developing the research hypothesis and interpreting empirical findings.

Argument 1 and Argument 2 show the commitment to the neoclassical theory. First, using rational choice theory as a benchmark makes it possible to identify behaviours that do not conform to what is expected and accordingly describe them as biases (A1). Likewise, consistently using the neoclassical standard also makes it possible to identify how many times and in which contexts these deviations take place (A2). In that sense, the idea that systematic flaws in human decision-making affect behaviour means accepting expected utility theory as the normative standard of decision-making and, accordingly, considering consistent deviations for this standard as systematic and predictable irrational biases. Argument 3 is also profoundly shaped by the neoclassical theory. The status of rationality failures requires having a preconceived notion of which outcomes are good or bad. From a neoclassical standard point of view, if a decision maximises utility, it is rational (and good). If it does not, it is irrational (and bad). Accordingly, within behavioural economics, decisions are weighted as proper (or improper), if they comply (or not) with rational choice theory expectations.

Nevertheless, rational choice theory status within behavioural economics poses some problems. The neoclassical model of behaviour has been highly contested. Numerous empirical and experimental findings, both in the lab and on field, call into question its predictive and descriptive power. Likewise, several authors question both its value as a normative theory of decision-making and its value as a benchmark for guiding research on decision-making (Elqayam & Evans, 2011; Gigerenzer, 2015). Therefore, while behavioural economics provides essential insights on how people react and make choices in everyday situations, its commitment to neoclassical theory raises doubts about some of its key findings and claims.

Notoriously, proponents of the fast-and-frugal heuristics approach (e.g. Gigerenzer, 2015; Grüne-Yanoff & Hertwig, 2016) argue that behavioural economics has become too eager to study deviations from rational choice theory and ignores the fact that people can take smart and intuitive decisions with positive results in contexts of uncertainty and with cognitive constraints. The approach questions some of the main behavioural economics related truisms, particularly the evidence that supports the notion that individuals suffer from systematic and unavoidable behavioural biases.

The limitations of behavioural economics in explaining behaviour represents a direct challenge to core arguments in the original nudge characterisation. If, due to the influence of rational choice theory, behavioural economics overstates the notion that people exhibit regular cognitive flaws, this poses problems with Argument 1 and Argument 2 and undermines the pragmatic value of nudges. Nudges are

supposed to work because agents are affected by behavioural biases and tend to behave like Humans rather than Econs. Likewise, biases must be general, systematic and predictable to ensure that nudges work in different settings. If this is not the case, nudges would no longer comply with property (iii), and their nature as behavioural change interventions would need to be revised. Likewise, if rational choice theory maintains its status as the normative model of behaviour, the judgement over what is rational carries its assumptions and limitations. This poses problems in terms of sustaining Argument 3 and indicates that behavioural economics might have limitations when it comes to supporting normative recommendations for policymaking and challenges the capacity of nudges to comply with property (iv).

Argument 4 also faces important objections. Even when accepting the empirical claims and main findings related to behavioural economics, critics maintain that the evidence is not strong enough to support the idea that people agree with the direction in which they are being nudged and would make the same choice themselves if it were not for behavioural biases. Many authors question the ability of nudges to improve individual wellbeing in line with people's preferences. Guala & Mittone (2015) recognise practical limitations when it comes to identifying and promoting people's true rational preferences through policy interventions. Sugden (2017) claims that empirical findings cannot support the argument and, instead, that it is based on the theoretical assumption that people have *rational latent preferences*. Similarly, Rebonato (2014) argues that to defend Argument 4, Thaler and Sunstein omit the criteria of revealed preferences and, instead, rely on the neoclassical assumption of rationality. White (2013) also questions behavioural economics and argues that it relies on a narrow *preference-constraints* framework, which is incomplete and insufficient for policymakers to identify people's *true* interests. Bovens (2009) also posits that *behavioural biases* are not the only factor that can produce a change in behaviour.

Many authors agree that, in practice, it is difficult to distinguish whether people make choices due to changes in preference or due to the influence of biases. Therefore, while some choices might be affected by behavioural biases and prompt individuals to act against their rational judgment, these cases might not be as frequent as Thaler and Sunstein assume (Bovens, 2009; Mitchell, 2004; Sugden, 2017). All of these arguments suggest that nudges struggle to uphold property (iv).

To sum up, Thaler and Sunstein's original nudge characterisation relies on a specific notion of human rationality. The evidence and claims from behavioural economics are vital to sustaining the empirical and normative value of the approach. However, considering alternative understandings of human rationality and revising the main findings of behavioural economics weakens some of the main arguments that support the original nudge characterisation. While some evidence supports Arguments 1, 2, 3 and 4, critics also argue that behavioural economics has significant limitations as a research program. These objections directly affect the value of nudges as a behavioural change strategy by questioning its empirical and normative value. Likewise, the objections indicate that some of the

supposed properties of nudges are not always met, which implies that the objection also poses problems of conceptualisation. Nudges fail to uphold properties (iii) and (iv).

### **3.2. Problems with the ways in which nudges steer choices.**

The implementation of nudges requires accepting a specific idea of how nudges affect choices. Thaler and Sunstein define nudges as interventions that (i) do not significantly change economic incentives, and (ii) do not forbid any options. Based on these two properties, the authors argue that nudges uphold property (v) and respect freedom of choice. This statement is vital to supporting the ethical value of nudges, based on the following argument:

(A5) Not reducing the number of choices and not altering economic incentives preserves freedom of choice.

Thaler and Sunstein defend Argument 5. According to the authors, because nudges do not change economic incentives and do not ban or exclude any options, they enable those who want to opt out of the nudge to do so easily and, therefore, are not coercive (Mills, 2013; Quigley, 2013; Sunstein, 2015c). However, most critics agree that not reducing the number of choices and not altering economic incentives are insufficient conditions for ensuring that nudges respect people's freedom.

Several authors argue that Thaler and Sunstein's claim that nudges respect freedom of choice is insufficient for capturing the possible ethical implications of nudges. The ethical value of nudges as a behavioural change strategy depends on how nudges influence behaviour. To be effective, nudges must exert some control or effect over people's behaviour and choices. Thaler and Sunstein omit this question and rely solely on Argument 5 to defend the benevolence of nudge as an intervention to influence choices. However, a closer look at the means that nudges employ to change people's behaviour suggests that the way in which nudges steer choices might be problematic.

Under property (iii), Thaler and Sunstein define nudges as an intervention that affects the behaviour of Humans, although Econs would ignore them. Property (iii) outlines why nudges work; it alludes to the fact that nudges change behaviour by exploiting or neutralising psychological mechanisms that are traditionally excluded from research on decision-making within economics. Thaler and Sunstein provide several examples of these mechanisms; however, they do not explicitly state by which specific mechanism the different nudges work; nor do they acknowledge how these mechanisms affect people's decision-making process.

Looking at property (iii), nudges have been criticised for not changing people's behaviour through rational persuasion and instead working via irrational mechanisms and taking advantage of people's cognitive flaws. Critics point towards a contradiction between property (iii) and property (v). Many authors suggest that nudges typically work covertly, without the decision-maker's awareness and bypassing deliberation. As a result, nudges face claims of being manipulative, non-transparent, and

being able to steer individuals towards a choice without their awareness or their consent. As a result, they are often described as a potential threat to agents' autonomy, agency and dignity. In line with these arguments, another important objection draws on both the short and long-term consequences of nudges. Detractors of nudges maintain that nudges do not help long-term decision-making. Since nudges do not foster deliberative engagement and override conscious reasoning, they create dependency and make people unable to make better choices in the future. Generally, these different objections stress that nudges are problematic due to the ways in which they steer choices. While Thaler and Sunstein defend that nudges automatically comply with property (v) due to properties (i) and (ii), several authors postulate that nudges often fail to comply with property (v).

#### **4. The need for a new understanding of nudges**

The original nudge characterisation defines nudges as an intervention that upholds five properties and that has both practical and normative value in terms of behavioural change. However, in the previous section, I explained that the approach faces a number of objections and counterarguments that pose problems for the original nudge characterisation. Objections to the underlying theory on human rationality and the ways in which nudges influence choices indicate that the arguments that support the original nudge characterisation have their limitations. Likewise, the objections make evident that nudges rarely uphold their supposed five properties and pure nudges might be less frequent than Thaler and Sunstein suggest.

Table 2 summarises the main counterarguments to the original nudge characterisation and the problems they create for the original understanding of nudges. The objections indicate two different things. Firstly, the concept of nudges that Thaler and Sunstein define faces objections about their value as a behavioural change intervention and their normative acceptability as a policy instrument for behavioural change. Secondly, the definition of nudges becomes problematic because very few nudges comply with the five properties and established in the original nudge characterisations.

**Table 2. Counterarguments against and problems related to the original nudge characterisation.**

Arguments	Counterarguments	Problems with the original nudge characterisation
(A1) People’s behaviour is frequently affected by behavioural biases.	Alternative understandings of rationality question the existence of behavioural biases.	It weakens property (iii) and the practical value of nudges. Without biases, nudges are not able to change behaviour.
(A2) Biases are systematic, predictable and generally unavoidable.	The influence of RCT on BE leads to BE overstating the cases in which biases affect decisions.	It weakens property (iii) and the practical value of nudges. If biases are not systematic, nudges are less useful when it comes to changing behaviour
(A3) Biases produce rationality failures and lead people to choices that undermine their wellbeing.	The influence of RCT on BE defines biased and unbiased choices. Alternative understandings of rationality question the notion of <i>rationality failures</i> .	It weakens property (iv). If there is no such thing as a rationality failure, nudges lose their normative value.
(A4) People have rational preferences	Evidence from BE is limited to supporting the idea that non-optimising choices are unwanted and always caused by biases.	It contradicts property (iv) and the normative value of nudges. Nudges do not always improve individual wellbeing as judged by the individuals themselves.
(A5) Not reducing the number of choices and not altering economic incentives preserves freedom of choice	The ways in which nudges steer choices are problematic, exploit irrational biases and compromise autonomy.	It contradicts property (v) and weakens nudges normative value. Nudges do not always respect freedom of choice.

The fact that nudges often fail to comply with all their properties leads to problems of conceptualisation. If we are trying to establish what a nudge is, and which interventions can be considered nudges, we can answer the questions in different ways. Some consider that the property that makes an intervention a nudge is that it respects freedom of choice. Certainly, Thaler and Sunstein put especial emphasis on this trait, so defining nudges using this condition appears appropriate. Others suggest that what distinguishes nudges from other types of behavioural change interventions is the mechanisms that they employ to change behaviour. In this sense, regardless of the ethical implications of the intervention, if an intervention changes behaviour by leveraging factors related to biases and blunders, temptation and following the herd, it should be considered a nudge. By contrast, in line with property (iv), others argue that nudges are interventions that correct rationality failures, regardless of the mechanism that they use or whether the nudge respects or infringes freedom of choice. Finally, property (iv) also suggest that nudges can be understood as an intervention that makes individuals better off according to their judgment, despite limiting freedom of choice.

In addition to the problems highlighted by the objections, a fundamental problem with the original nudge characterisation is that Thaler and Sunstein present a unitary idea of a nudge and assume that the properties assigned to nudges are automatically related to each other. The authors argue that, because

nudges do not change economic incentives (i) and do not forbid any options (ii), nudges respect freedom of choice (v). Likewise, they argue that because nudges use specific psychological mechanisms (iii), they always work in situations in which there is a potential rationality failure and always improve people wellbeing as judged by the individuals themselves (iv). The properties are not contradictory; they can be consistent with one another, and in particular cases, it is possible to find interventions that comply with all of them. However, it would be wrong to assume that the properties always go together. For instance, even though nudges do not change economic incentives and keep all the available options, due to the mechanisms that they use to change behaviour, they may compromise people's capacity to make their own choices. Likewise, even though nudges work via mechanisms that are traditionally regarded as non-rational, if they work, they may change rational and non-rational decisions. Finally, while nudges might correct rationality failures and steer people towards choices with a positive impact for their wellbeing, that does not imply that people generally agree with the direction in which they are being nudged. The missing link between the properties indicates that the concept of nudges needs to be revised.

In the following chapters, I argue that, beyond addressing the main arguments and counterarguments to support or reject the original nudge characterisation, it is necessary to rethink how nudges can contribute to policymaking. I will argue that, in light of the extensive discussion on nudges, the original nudge characterisation seems too ambitious. To work as intended, nudges ought to comply with properties (i), (ii), (iii), (iv) and (v). In reality, for a number of reasons, this is often implausible. The upcoming chapters discuss in depth some of the main objections to nudges, focusing on the ethical acceptability of nudges (Chapter 2) and behavioural insights used in public policy (Chapter 3). The thesis defends the use of nudges in policymaking and accordingly offers arguments that resolve some of the objections.

However, the project is not a defence of Thaler and Sunstein's understanding of nudges. As many authors claim, and as I will argue in the following chapters, the original nudge characterisation has significant problems and needs to be revised. While the project is critical of Thaler and Sunstein's idea of nudges, it is optimistic about the value of nudges within policymaking and argues that, despite the objections, nudges still have practical value as a policy instrument for behavioural change. I will argue that the main problem is the normative content. The original nudge characterisation assigns a substantial normative burden to nudges, something that is unusual when defining policy instruments. Normative properties (iv) and (v) are the source of the main problems and have monopolised the debate about their implications, advantages and limitations. In the following chapters, I will try to defend the pragmatic value of nudges for policymaking, revising what we can expect from their use if we drop the central normative claims and properties. I will also discuss and revise whether, when looking beyond the original characterisation, nudges can be both more acceptable and more useful when adopted to pursue behavioural change.

## Chapter 2

### The ethics of nudges

#### 1. Introduction

The use of nudges has triggered an interesting and significant philosophical debate. Their implementation has attracted much controversy, and there are significant ethical objections to their use. In general, their legitimisation as a valuable policy tool is still up for debate. In the original characterisation of nudges, Thaler and Sunstein assigned a firm moral intention to nudges. Firstly, they argued that nudges improve people's wellbeing, and that people agree with the direction in which they are being nudged. Secondly, they argued that nudges respect agents' freedom of choice. Finally, they remarked that due to the influence that the context has on choices nudges are unavoidable and, therefore, opposing their use is pointless.

Despite the overall positive intention of nudges, it has become apparent in the discussion that Thaler and Sunstein's arguments are unsatisfactory when it comes to supporting the moral acceptability of nudges. For the most part, critics find nudges problematic on many grounds. Several authors suggest that nudges fail to improve individual wellbeing as judged by the individuals themselves; therefore, their use cannot be justified under this claim. Likewise, controversy arises over that fact that while nudges maintain all the available options, the ways in which nudges steer choices fail to respect agents' freedom of choice and autonomy. Finally, critics maintain that, even if many uncontrolled factors affect behaviour, the use of nudges to steer people's behaviour is always intentional and requires justification. Overall, critics argue that nudges fail to comply with normative standards and carry added normative costs.

When these three points are examined, they raise important normative objections, and have prompted a number of arguments against nudges. While these arguments are diverse, typically they can be classified as (a) arguments against the aims of nudges, and (b) arguments against the means of nudges. By "the means of nudges" I mean the ways in which they steer people's choices.

Objections under category (a) — against the aims of nudges — comprise arguments that maintain that nudges are problematic due to the goals they promote. Critics argue that to support the idea that nudges promote people's wellbeing, Thaler and Sunstein make unfounded assumptions about people's preferences and ignore evidence to the contrary. Subsequently, critics doubt the legitimisation of nudges and accuse nudges of overreaching and covert paternalism (Gigerenzer, 2015; Infante et al., 2016; Rebonato, 2014; Sugden, 2017; White, 2013).

Objections under category (b) — the means of nudges — comprise arguments that object to how nudges steer people's choices. The mechanisms that nudges use to influence choices appear problematic

because they bypass deliberation, are unnoticeable and exploit irrational biases (Bovens, 2009; Hausman & Welch, 2010). Accordingly, nudges face objections of manipulation, lack of transparency, and adverse long-term effects and critics accuse nudges of compromising people's autonomy and agency. Within this discussion, nudges are deemed an illegitimate tool due to their features as an intervention aimed at behavioural change.

The debate on the ethics of nudging is so extensive that it seems to indicate that something is particularly wrong with nudges. While objections raise important points about remaining vigilant about the use of nudges in public policy, the discussion overstates the negative implications of nudges and overlooks elements that play in favour of legitimising them normatively as a policy instrument.

Chapter 2 explores the principal ethical objections to the use of nudges by considering both (a) the aims of nudges and (b) how nudges work. The chapter aims to map frequent objections to the dimensions of nudges and address their implications and limitations with the aim of drawing normative conclusions about nudges. Below, I argue that nudges are not intrinsically problematic as a result of either their aims or their means. I suggest that normative considerations about their use in policymaking require that their ethical implications be examined beyond the libertarian paternalism framework, by exploring the way in which different nudges affect behaviour and taking into account empirically grounded approaches to decision-making.

The chapter proceeds as follows. Section 2 looks at the debate about the aims of nudges. It provides an overview of the main arguments against nudges and responds to the main objections. The discussion argues that the aims of nudges are not pre-established and can be in accordance with the promotion of the goals and outcomes that governments seek to achieve through other behavioural strategies. Section 3 addresses the objections about the means of nudges, particularly the claim that nudges undermine individual autonomy. The discussion identifies two main objections: the rationality objection and the reflection objection. While both contain legitimate worries, there are two different problems. Firstly, critics discuss the ethical value of nudges using a conceptually vague characterisation of nudges and overlooking the differences between interventions. Secondly, critics base their objections on ideal notions of autonomy and decision-making that lack empirical bases. Responding to these two objections, section 3 argues that the means of nudges are not intrinsically problematic in terms of autonomy. The last section discusses possible normative standards for evaluating nudges. It argues that the moral acceptability of nudges as a policy instrument is a combination of three aspects; the type of nudge, the context of use and the aim it promotes. Finally, the chapter ends with some conclusions.

## **2. Objections to the aims of nudges**

The first argument in defence of the ethical acceptability nudges is the claim that nudges improve people's subjective wellbeing. Thaler and Sunstein argue that nudges attempt to influence choices to 'make the choosers better off as judged by themselves' (Thaler & Sunstein, 2008, p. 5). Drawing on

findings from behavioural economics, Thaler and Sunstein argue that nudges improve individual wellbeing and promote choices that agents would choose if their decisions were not affected by cognitive biases, poor self-control and limited time, information and cognitive abilities. Given these factors, Thaler and Sunstein maintain that policymakers ought to interfere in people's choices and steer them towards the options that their rational self would have chosen free from the influence of decision biases. The argument is vital to defending the ethical value of nudges; the implementation of nudges is deemed ethically legitimate because nudges promote wellbeing, and people agree with the direction in which they are being nudged. However, the general discussion around nudges indicates that this first justification is problematic. Many authors argue that nudges often fail to promote people's subjective wellbeing and lack legitimacy.

### **2.1. Nudges and wellbeing, as judged by individuals themselves?**

Thaler and Sunstein claim that nudges steer individual wellbeing as judged by the individuals themselves. As mentioned in Chapter 1, they rely on the following four arguments to support this claim:

- (A1) People's behaviour is frequently affected by behavioural biases.
- (A2) Biases are systematic, predictable and generally unavoidable.
- (A3) Biases produce rationality failures and lead people towards choices that undermine their wellbeing.
- (A4) People have rational preferences and prefer to avoid the effect of behavioural biases.

These arguments are partly supported with empirical evidence, particularly by findings from behavioural economics. However, they face significant criticism and epistemic and methodological challenges.

An active line of criticism regarding nudging has focused on assessing argument four. The central theme of argument four is that people are susceptible to decision biases that prevent them from choosing based on their true rational preferences and their choices no longer reflect what they judge to be good for themselves. Thaler and Sunstein argue that external authorities should respond by interfering in people's choices and nudging them towards options that promote their wellbeing.

Infante et al. (2016) and Sugden (2017) argue that Thaler and Sunstein lack subjective information about what people want, and adequate criteria to prove that people's decisions systematically fail to reflect their true preferences. According to Infante et al. (2016) normative behavioural economics uses a model of preference purification that assumes that 'an inner rational agent is trapped inside a psychological shell' (Infante et al., 2016, p. 2). When choices are inconsistent with rational choice theory, the assumption is that agents have made a mistake due to the effect of decision biases. In this context, policymakers need to change people's choices and reconstruct their true preferences. However,

the model does not delve into the psychology of choices and only works on the assumption that people's choices do not reflect their preferences. According to Thaler and Sunstein, it is evident that people prefer to be fit and healthy than overweight and unhealthy, and it is evident that people prefer to cut senseless expenses in the interests of a wealthier future. However, Thaler and Sunstein fail to provide evidence that these inferred preferences actually reflect people's true preferences. The existence of rational preferences is assumed but is not empirically proven (Sugden, 2017, p. 117).

Sugden (2017) argues that Thaler and Sunstein do not provide satisfactory criteria to distinguish choices that are the result of a cognitive bias from decisions motivated by alternative factors. To establish whether a decision is good or bad, Thaler and Sunstein only take its outcome into consideration: if the choice maximises utility, it is rational; if it does not, it is a mistake resulting from a cognitive bias. However, their approach fails to provide appropriate evidence that the choice is an error of judgment. In this regard, Thaler and Sunstein lack empirical evidence about people's true preferences and do not have criteria to identify which choices are affected by cognitive biases.

To illustrate these points, we will use the example of Peter and John. Peter and John are both 35 and work at the same company; however, they have very different preferences when it comes to food and health. Peter likes to stay fit and, with some effort, he can maintain a proper diet. John likes to eat whatever he wants. Peter strives to live a long and vigorous life. John prefers to enjoy the present, even at the expense of his life expectancy. During the week, both Peter and John have lunch at their company's canteen and often choose the chocolate cake for dessert. Because we know how they feel about food, health and life, we can tell that Peter's decision goes against his rational preferences, while John's does not. Consequently, we can assume that Peter would agree with a hypothetical nudge while John would not.

Thaler and Sunstein would interpret this differently. Even though John and Peter's motivations differ, Thaler and Sunstein would not be able to distinguish how. According to their criterion, because both Peter and John chose against their long-term health, against the standard rational-optimising option, both choices would be considered irrational. By Thaler and Sunstein's standards, both Peter and John are choosing against their true rational preferences; hence, both of them would benefit from the nudge, and both would agree with its aim. However, as Sugden (2017) argues, and as the examples illustrate, Thaler and Sunstein ignore what Peter and John wanted. By only looking at the outcome, Thaler and Sunstein can easily misconstrue the cause of choice, assuming an error of judgment when there is not one.

In general, while some decisions might appear irrational, they can be consistent with the individuals' overall preferences. Thaler and Sunstein do not have real information about people's preferences and do not provide criteria to identify choices affected by cognitive biases. It is only by assuming a set of

hypothetical rational preferences that they can claim that nudges promote people's subjective wellbeing. Therefore, argument four can only be maintained under specific assumptions on people's preferences.

Along similar lines, Rebonato (2014) identifies a methodological problem with argument four. To support argument four, Thaler and Sunstein discard revealed preference theory. According to Thaler and Sunstein, evidence from behavioural economics indicates that people's choices do not reflect their real intentions and rational preferences. However, their model does not provide any empirical criteria to identify these cases. Neoclassical economics commonly uses revealed preference theory to identify what people want and prefer. Behavioural economics reveals essential limitations with the theory but does not provide alternative criteria to identify in which situations choices reflect preferences and in which situations they do not. Under the criterion of revealed preferences, the policy implications about when and why it is necessary to interfere in people's choices differ significantly.

Going back to the example of Peter and John, advocates of revealed preference theory would argue that, in order to understand what Peter and John prefer, we should look at what they do. In this case, because both Peter and John choose the cake, we should assume that they both prefer it. By this interpretation, a nudge would not be necessary; neither Peter nor John would want to be nudged towards any other option. Because we know how they really feel, we can determinate that, in this scenario, John wins and Peter loses. However, with the criteria of revealed preferences, the decision would be not to interfere.

In the Peter and John situation, a hypothetical nudge would improve Peter's wellbeing but compromise John's; the lack of a nudge would benefit John but impair Peter. Regardless of how we choose to proceed, the point is that the evidence to support Thaler and Sunstein's claim is not incontestable. Infante et al. (2016), Rebonato (2014) and Sugden (2017) identify empirical and methodological problems with argument four which indicate that in practice, policymakers face problems in identifying people's *true* preferences. These arguments translate into problems with the normative foundations of nudging. If the evidence to support the fact that people have rational latent preferences is insufficient, and we lack standards to identify in which situations decision biases compromise people's choices, it is unclear whether nudges can generally improve individual wellbeing as judged by the individuals themselves.

From an alternative understanding of human rationality, Gigerenzer (2015) questions the evidence underpinning arguments two and three. According to Gigerenzer, the idea that people make regular errors of judgment exaggerates people's cognitive limitations by uncritically accepting the rational choice approach (Gigerenzer, 2018). Gigerenzer argues that behavioural economics lacks empirical evidence to support the notion that people predictably and systematically lack rationality and remarks that decision-making mistakes due to cognitive biases are not as prevalent as Thaler and Sunstein assume. Accordingly, he questions the use of nudges under the justification that they promote people's true preferences. Gigerenzer claims that embracing the standard justification of nudges blames people

for their own mistakes, while failing to recognise that, in general, their decisions are correct. In this sense, using nudges under the standard justification can lead policymakers to refrain from considering alternative strategies for behavioural change. Gigerenzer argues that, beyond nudges, behavioural evidence supports the use of interventions that intend to educate people, improve their deliberative capacities or correct external factors that may be affecting their decisions (Gigerenzer, 2015, 2018).

In a very confrontational critique, White (2013) argues that behavioural economics has essential limitations when it comes to supporting the claim that nudges promote individual subjective wellbeing, because people have different understandings of what is good for them. Given these limitations, White argues that nudges should be rejected because, under the original justification, nudges entail value substitution that replaces agents' judgment of what is best for them, with the policymaker's interest over what they should be doing. This value substitution is ethically problematic because it gives policymakers *carte blanche* to nudge as they please, without actually justifying when and why nudges promote positive goals.

Under libertarian paternalism, nudges are legitimate when they promote people's subjective wellbeing. The above-mentioned arguments assert that this cannot always be ensured using Thaler and Sunstein's method. With the current criteria, Thaler and Sunstein overemphasize the instances in which nudges can promote people's preferences, revealing an imbalance between what Thaler and Sunstein claim nudges do and what nudges can actually do. An obvious solution to this problem would be to ask people what they think about the nudge's aim. If they accept it, the nudge could still be legitimate. However, this option has obvious practical limitations to legitimising nudges in policymaking. Asking individuals how each of them feels about diverse topics would be a very costly and time-consuming process. Likewise, we can expect a considerable heterogeneity of opinions depending on people's views on different issues (Guala & Mittone, 2015; White, 2013). Asking people how they feel about the aims of specific nudges also contradicts Thaler and Sunstein's main argument. Hausman (2018) contends that if we doubt that people's choices represent their true preferences in the first place, there is no reason as to why we should accept their own opinions over what is right for them. Their answers might also reflect preferences shaped by psychological limitations; thus, it is unclear why they should be considered a reliable source. Therefore, knowing individuals' preferences may not be the most practical option for useful public nudging.

Problems with arguments two, three, and four pose challenges to nudges' capacity to improve individual wellbeing as judged by the individuals themselves. Alternative research programmes on rationality and competing criteria to identify people's preferences indicate that the idea that people always agree with the aims of nudges is based on assumptions about human behaviour and rationality rather than on substantial empirical evidence about individuals' preferences. These arguments pose problems for the ethical acceptability of nudges. Under this justification, nudges appear to be problematic because policymakers might use this claim to conceal an excessive paternalistic agenda or to promote illicit

ends. Likewise, the approach solely blames individuals for their decision-making mistakes and may discard the use of other interventions with potentially better performance. In conclusion, these arguments suggest that if nudges fail to promote subjective wellbeing, their ethical value as a policy instrument is in doubt.

## **2.2. Nudges, alternative aims, and justifications.**

When discussing the problems nudges have with fulfilling their original goal, many authors have explored nudges' potential to promote different aims and how their use is ethically acceptable under diverse normative justifications.

Carrying on with the Thaler and Sunstein's original intention, many authors have discussed nudge's potential to improve people's wellbeing, regardless of their preferences. Often there can be inconsistencies between what people want and what is right for them. In these situations, it is common for governments to interfere in people's choices and justify this interference under paternalism. Paternalism can be defined, as 'the interference by some outside agent in a person's freedom for the latter's good' (Le Grand & New, 2015, p. 7). This includes cases in which governments interfere in people's choices to promote particular outcomes, even at the expense of agents' preferences. For instance, governments require people to wear seatbelts in cars and helmets when riding motorcycles, prevent people from swimming or skiing in potentially harmful weather conditions and ban drugs to prevent their adverse effects. In line with these interventions, nudges which seek to make people engage in healthy habits or save up money for the future, for instance, seem to follow a paternalistic intention. Indeed, most of the examples illustrated by Thaler and Sunstein seek to help people to get healthier and wealthier by assuming that this is what they want; therefore, policymakers could defend their use under a paternalistic justification and argue that 'nudged individuals are always better off independently of their preferences' (Guala & Mittone, 2015, p. 386)

Behavioural findings provide new insight into the ways and situations in which people fail to make the right decision for themselves. At the same time, these findings raise questions about how policymakers should address them and in which situations it is legitimate and positive to interfere in people's choices. The policy and welfare implications of behavioural economics are an unresolved issue that includes diverse and conflicting views. Thaler and Sunstein advocate for soft, liberty-preserving nudges. They argue in favour of interfering and correcting failures of rationality but still allowing people to choose. Conly (2013) contends that the negative consequences of cognitive biases and the limitation to overcoming them through education require strong government intervention. While sharing Thaler and Sunstein's diagnosis, Conly (2013) also points out that nudges might be ineffective when it comes to ensuring actual and lasting behavioural change and argues in favour of using more coercive interventions. By contrast, under the *opportunity criterion*, Sugden (2004) suggests that this evidence does not automatically justify intervention. The debate around the normative base of welfare economics

is profound, yet regardless of how we approach this topic, what is relevant for the use of nudges is that, if we choose to interfere, nudges have a practical value in promoting wellbeing and their use is justifiable from a paternalistic standpoint.

Beyond individual wellbeing, some authors have considered the advantages of using nudges to tackle social problems. Nudges can be useful in resolving issues where the need for intervention is already justified by traditional economic grounds, for instance, in cases of externalities, public goods and information asymmetry (Chetty, 2015; Loewenstein & Chater, 2017). For instance, Guala & Mittone (2015) argue in favour of using nudges to solve public policy problems, particularly to correct externalities. The authors highlight the fact that Thaler and Sunstein had already considered this option themselves in *'Nudge'*, when they mention the potential of nudges to promote energy conservation, organ donation and tax compliance.

Similarly, Loewenstein & Chater (2017) point out that incorporating insight on nudges into issues of public policy can be highly beneficial when it comes to rethinking the tools available for changing behaviour in situations of coordination or social dilemmas. For instance, on matters concerning public goods, the traditional economic framework asserts that it is rational for self-interested individuals to act as free riders. Drawing on these assumptions, interventions aimed at correcting this behaviour rely exclusively on incentives, bans and the privatisation of public goods (Guala & Mittone, 2015). Behavioural and social sciences have shown that there could be biased motivations underlying what economics classifies as free rider behaviour. For instance, cognitive biases could be the reason why people fail to cooperate, due to inertia, social influence or the effect of elements of the choice architecture. If this is the case, a nudge could be very useful in improving cooperation and reducing externalities. On a similar note, Nagatsu (2015b) describes the concept of *social nudges* and evaluates the potential of nudges to improve the provision of public goods. In relation to public goods, extensive evidence in social science suggests that individuals are conditional co-operators rather than free riders, meaning that they have preferences for cooperating if others also cooperate (Bicchieri, 2005; Elster, 2007). In this case, there is no bias; however, using a nudge that communicates the behaviour of others may activate a social norm and could be helpful when it comes to increasing cooperation. These examples illustrate that nudges have the potential to pursue different policy goals and can complement traditional policy interventions.

Using nudges to tackle any policy problem opens up the possibilities of how to legitimise their use. Commenting on alternative ways to justify the implementation of nudges, Kelly (2013, p. 213) maintains that the same techniques that Thaler and Sunstein suggest using to make individuals better off can also be utilised to promote utilitarian and Rawlsian goals. The use of nudges may undermine individual freedom of choice and may involve paternalism; however, their use may still be legitimate if they improve overall social welfare, reduce inequalities or ensure access to primary goods.

In the original characterisation of nudges, Thaler and Sunstein outline the rationale for intervening with nudges (the existence of failures of rationality) and define the ethical legitimisation of nudges with libertarian paternalism (promoting individual subjective wellbeing). Section 2.1 argues that existing evidence falls short of supporting the argument for the ethical acceptability of nudges. If people have to agree with their aims for nudges to be legitimised, their implementation would be much more restricted than Thaler and Sunstein contend. Likewise, if policymakers were to use nudges under the original justification, their use would be problematic and would raise relevant normative objections.

However, we should not reject nudges simply because they fail to promote people's subjective wellbeing. When discussing the aims of nudges, inevitably they are presumed to be linked to libertarian paternalism. In fact, in previous literature, both terms are often confused and used interchangeably (Gigerenzer, 2015; Hansen & Jespersen, 2013; Schubert, 2015). However, nudges and libertarian paternalism have distinct and separable meanings. Libertarian paternalism is a normative framework that aims to promote specific values and ends; nudges are tools to influence behaviour and encourage behavioural change. Nudges should be understood as a policy tool, a tool with practical value for promoting pro-self and pro-social goals, the implementation of which can respond to different normative justifications.

White (2013, p. 83) argues that nudging is not about helping people make better choices, but about getting people to make the choices that 'policymakers want them to make'. The point is that this is not necessarily a problem, particularly if nudges promote ethically consistent goals. It is clear that resolving current public policy problems requires a change in people's behaviour, and that traditional government tools sometimes fail to tackle such issues. Nudges working complementarily with other tools can shape individual behaviour to match different objectives and motivate behavioural change. As far as their aims are concerned, nudges are not inherently problematic.

This section has attempted to clarify the objection against nudges based on the aims that nudges promote. It has argued that, even though nudges have limitations when it comes to improving individual wellbeing as judged by individuals themselves, they retain practical value when it comes to supporting both pro-self and pro-social goals and their use can be justified under different normative bases. Nudges should be understood as policy instruments and, as such, should be used to pursue relevant and ethically underpinned goals, as well as aims that have to be publicly discussed and accepted by the majority. Essential questions remain about whether it is legitimate to employ nudges. To assess the ethical value of nudges, one has to consider if their use responds to relevant and legitimate goals and ensure that nudges are done under democratic control. However, these questions do not apply to nudges alone.

Addressing objections to the aims of nudges responds to some of the objections against nudges. However, ethical concerns about implementing nudges remain. Several ethical issues regarding nudges are overlooked if they are viewed as 'just another mode of influence in the toolbox of state controls'

(Baldwin, 2014, p. 832). Nudges face significant objections about the way in which they steer people choices; the following section tackles this question.

### **3. Objections to the means of nudges**

The second argument in defence of the ethical acceptability of nudges is the claim that nudges respect agents' freedom of choice. Thaler and Sunstein argue that, in general, people should be free to choose what they want to do according to their preferences. For this reason, libertarian paternalism attempts to design policies 'that maintain or increase freedom of choice' (Thaler & Sunstein, 2008, p. 5). As outlined in Chapter 1, Thaler and Sunstein maintain that nudges are not coercive. They argue that, because nudges do not change economic incentives and do not ban or exclude any options, they enable those who want an easy way to opt out of the nudge. In this sense, they emphasise that nudges are liberty preserving and 'make it easy for people to go their own way' (Thaler & Sunstein, 2008, p. 5). Nudges guide people towards better outcomes, but never force them to do anything they do not want to do. The original characterisation of nudges defines them as interventions that respect freedom of choice (property v. of the original characterisation of nudges) and this feature supports their implementation as a behavioural change strategy.

Nonetheless, several authors note that the fact that nudges do not enforce explicit barriers to liberty is insufficient to address their ethical implications. The majority of critics agree that even though nudges do not block choices, do not modify economic incentives and typically maintain all the available options in a context of choice, they interfere in people's decision-making and diminish their ability to make their own choices. Many found nudges problematic due to how they steer people's behaviour; specifically, they express concerns about nudges' effect on individual autonomy (Hansen & Jespersen, 2013; Hausman & Welch, 2010; Kelly, 2013).

Blumenthal-Barby, 2013 states:

'The crucial point is that even if choice architecture does not block or significantly burden choices, it might still interfere with a person's ability to discern and consider options and act according to her own preferences, i.e., it might interfere with her autonomy' (p. 190).

She further argues that if nudges do not engage rational argument and instead bypass reasoning by influencing choices in a way that is not obvious, then they constitute an act of manipulation that also undermines individual autonomy.

Critics argue that, whatever the goal, intention or result of a nudge, nudges are wrong due to the way in which they operate. In this sense, how nudges change behaviour is perceived as inherently problematic. The essential idea is that nudges work through mechanisms that compromise people's capacity to make their choices according to their own preferences and desires.

Despite general claims that nudges compromise autonomy, the objections reflect different dimensions and understandings of autonomy. The debate about the problematic nature of nudges is fundamentally shaped by the conceptions of autonomy, rationality and decision-making used to discuss their implications. The debate is complex because it includes normative and descriptive ideas about how individual preferences are formed and should be formed, which factors affect and should affect decision-making, and which elements promote or undermine personal autonomy.

The discussion on autonomy is particularly complicated because autonomy is an abstract and strongly disputed concept with different meanings. Broadly speaking, autonomy refers to the idea of a person being self-determining, a person whose decisions and actions are their own, who maintains the capacity to reflect upon one's preferences, reasons and actions, and can make their own choices free from external interventions (Dworkin, 1988) While several conceptions of autonomy seem to share this general intuition, what it means to be self-determining and in control of one's own decisions and actions varies significantly depending on the conception of autonomy at stake.

Despite the many worries and different conceptualisations of autonomy and decision-making discussed within nudge literature, objections that find nudges problematic due to how they steer choices tend to focus on two ideas:

- a. Nudges are problematic because they trigger non-rational psychological mechanisms.
- b. Nudges are problematic because they impede or obstruct reflection.

Some critics note that nudges threaten autonomy because they work via irrational mechanisms and take advantage of people's cognitive flaws (Bovens, 2009; Conly, 2013; Hausman, 2018; Hausman & Welch, 2010; White, 2013). Many authors indicate that working through non-rational mechanisms compromises people's autonomy. Other authors suggest that nudges typically work covertly without decision-makers being aware of the nudge and by bypassing or obscuring deliberation (Bovens, 2009; Grüne-Yanoff & Hertwig, 2016), leading to these factors undermining reflectiveness and compromising people's autonomy.

In this first theme, the focus is on rational deliberation, and the fear is that nudges trigger non-rational psychological mechanisms. I use the expression "the rationality objection" to represent these worries and objections. In the second theme, the focus is on reflection and conscious deliberation, and nudges appear problematic because they impede or obstruct it. I use the expression "the reflection objection" to refer to these objections. In previous literature, there is a significant overlap between the arguments; however, differentiating the cases seems to be the best way to address which elements of nudges are problematic.

Below, I discuss these two objections. I analyse why nudges appear problematic, and under which arguments and assumptions those objections hold true. Firstly, I note that in both cases, critics tend to misrepresent nudges. Critics often group nudge interventions into the same category, relying on

elements that are difficult to conceptualise and not shared by all nudges. As a result, they assess them as a general category and, therefore, fail to take the differences between nudges into consideration. Secondly, I argue that idealistic understandings of decision-making and autonomy underpin both objections. Critics employ conceptions of rationality and reflection that lack psychological insights and rely on assumptions unsupported by the empirical evidence on decision-making.

### **3.1. The rationality objection**

The rationality objection states that nudges are problematic because they trigger or take advantage of irrational psychological mechanisms. Hausman and Welch (2010, p. 130) argue that nudges play on ‘flaws in human judgment and decision-making to shape people’s choices’. Similarly, Bovens (2009, p. 209) argues that what distinguishes nudges from other types of influences in shaping choices is the fact that ‘some pattern of irrationality is being exploited’. On the same note, Grüne-Yanoff and Hertwig (2016, p. 153) state that ‘what is genuinely novel about the nudging approach [...] is the idea of exploiting people’s cognitive and motivational deficiencies’. Conly (2013, p. 30) also remarks that, when nudging, ‘rather than regarding people as generally capable of making good choices, we outmanoeuvre them by appealing to their irrationality, just in more fruitful ways’.

In general, authors agree with the fact that influencing choices by triggering irrational responses is ethically problematic. According to Hausman and Welch (2010, p. 128), when shaping choices ‘does not take the form of rational persuasion, autonomy—the extent to which individuals have control over their own evaluation and deliberation—is diminished’. On a similar note, (Bovens, 2009, p. 209) points out that ‘there is something less than fully autonomous about the patterns of decision-making that *Nudge* taps into’ and argues that ‘when we are subject to the mechanisms that are studied in “the science of choice”, then we are not fully in control of our actions’.

The above citations suggest that the problem with nudges is the fact that they exploit irrationality. For example, framing devices that highlight losses appear harmful because they affect behaviour by exploiting loss aversion and, therefore, change behaviour by exploiting an irrational bias without involving rational reasons. Similarly, default rules appear to be problematic because they do not constitute a rational reason to change preferences over options (Bovens, 2009). Likewise, the communication of social norms works because it triggers an irrational response and is problematic because agents do not make choices based on a rational consideration (Bovens 2009; Hausman & Welch 2010; Schubert 2015). In the rationality objection, the essential idea is that exploiting irrationality compromises people’s autonomy.

The objection to nudges on the basis that they exploit irrationality suffers from two main problems. Firstly, not all nudges exploit irrationality, and stipulating which interventions do and which do not appears to be somewhat challenging. Secondly, critics rely on rational choice theory as the normative

foundation of behaviour, a framework with notorious problems, to argue that a) nudges exploit irrationality and b) exploiting irrationality compromises people autonomy.

### *3.1.1. Nudges and rationality*

Let us start by considering the relationship between nudges and rationality. The most common answer for defending nudges against the rationality objection is to argue that not all nudges exploit irrationality. Sunstein (2015b), for instance, responds to critics by claiming that many nudges, such as, education campaigns, informational campaigns, reminders, warnings and the provision of feedback are interventions that engage rational deliberation and do not constitute a threat to people's rationality. However, there is a problem of conceptualisation within this debate. It is difficult to articulate whether these interventions should count as nudges because there are no precise definitions of a nudge. The border between which interventions are nudges and which interventions are not is diffuse. Several interventions count as nudges because they do not change economic incentives and do not ban or exclude any option; however, it is far less clear how they comply with properties iii, vi and v. Sunstein employs a broad definition, which fits many different types of interventions that do respect rational deliberation. If one adopts his definition, then indeed many nudges respect rationality. However, even if we adopt Sunstein's approach, the fact that some so-called "nudges" can be considered generally unproblematic does not resolve concerns about those interventions that work by exploiting cognitive flaws.

In recent publications, critics seem to direct the rationality objection only at those interventions that rely on people's cognitive biases. For instance, Hausman (2018) uses the term "nudging" to reference 'changing the choice circumstance to neutralise or to exploit deliberative foibles' and explicitly distinguishes nudges from other ways of steering choices such as information, education or deceiving. However, even when more refined, it is challenging to articulate which decision-making factors neutralise or exploit irrational biases.

As mentioned in Chapter 1, Thaler & Sunstein (2008, p. 37) stress the idea that human behaviour is "nudgeable", which means it can be easily influenced through mechanisms not considered within the traditional economics framework. The authors categorise factors that affect Humans and not Econs into three groups: biases and blunders, temptation, and following the herd (Mongin & Cozic, 2018; Thaler & Sunstein, 2008)

Broadly, by "biases and blunders" they are referring to the influence of heuristics such as representativeness, anchoring and adjustment, and availability; and biases such as overconfidence, loss aversion, and framing effects (Tversky & Kahneman 1974; Kahneman & Tversky 1979). By "temptation", they are referring to a lack of will power and general failures to control impulses and maintain self-control. By "following the herd", they emphasise the significance of social influence and social norms in shaping agents' decision-making. Within behavioural economics, these factors are

conceived as deviations from rational choice and regarded as non-rational. Designing choice architecture in ways that engage with these psychological factors implies that people's choices are affected by irrational factors. However, whether these factors are irrational is not a straightforward point. The factors described in '*Nudge*' as decision-making flaws do not hold true if interpreted according to alternative understandings of rationality (Gigerenzer 2015; Schubert 2015; Mongin & Cozic 2018).

Firstly, let us start with the "biases and blunders" category. The "biases and blunders" category references factors such as framing effects, loss aversion and other psychological biases found under the heuristics and biases research programme. Within behavioural economics, framing effects and loss aversion are presented as typical examples of decision-making biases. They occur when an agent's preferences between two identical sets of alternatives vary depending on how options are described, particularly if they are described as losses rather than gains. These findings violate the axioms of the expected utility model and are accordingly classified as biases. However, adopting an alternative approach, Gigerenzer (2008) maintains that, in cases where two definitions of the same situation are framed, they are logically equivalent but not informationally equal. Given contextual and cognitive constraints, agents use *fast and frugal heuristics* for breaking down pertinent information and decide between options. In these situations, decisions shaped by framings effects are not irrational but *ecologically rational*, i.e., rational in a particular context. The fast and frugal research programme generally challenges the findings of the heuristics and biases research programme. Research on ecological rationality establishes that many supposed irrational biases are "ecologically rational" across environments, given conditions of uncertainty and limited cognitive resources.

The "temptation" category, references issues of self-control, failures to resist temptation and time inconsistency. According to different conceptions of rationality, these factors often qualify as failures of rationality. Decision-making research on temporal choice has verified a systematic preference for small but imminent rewards over greater but delayed rewards. Agents discount future utility, a phenomenon labelled as *temporal discounting*. In behavioural economics, research on temporal choices follows the *hyperbolic discounting model*, a model that asserts that when agents face a choice between an inferior early option and a superior later option, they prefer the latter when both options are remote in time but switch to preferring the former as the time for both options approaches. The change of preferences appears irrational because agents' preferences are not consistent over time. However, previous literature indicates that rational factors could also explain choice inconsistency and preference reversal. Some research indicates that inconsistent changes in preferences over time can result from imperfect foresight, lack of information, or recently learned information, factors that might induce individuals to change or update their preferences (Gabaix & Laibson, 2017). Therefore, while inconsistent time preferences often reflect irrationality, it is empirically difficult to identify in which situations this is the case and to distinguish these cases from rational cases of preference reversal.

In terms of “following the herd”, decision-making research provides extensive evidence as to why following a social norm or following the behaviour of others is unlikely to be irrational. According to broader notions of rational choice theory that question the assumption of unbounded self-interest, game theory research suggests that as far as public goods are concerned, individuals have conditional preferences for conformity meaning that they prefer cooperation if a sufficient number of others also cooperate (Bicchieri 2006; Elster 2007). Nagatsu (2015b) maintains that nudges that provide information about the behaviour of the majority mobilise these preferences and trigger a rational response. Gigerenzer (2015) also contends that observing and following other’s behaviour in the context of uncertainty and limited information is following a social heuristic, which is an “ecologically rational” strategy. Likewise, Hedström (2006) argues that observing and copying other behaviours in cases of limited information is a rational strategy, a mechanism usually called “rational imitation”.

The examples described highlight alternative interpretations of the same phenomena as rational or irrational, depending on the understanding of rationality employed<sup>1</sup>. Some nudges are designed to engage with psychological mechanisms traditionally regarded as non-rational within economics. However, the categorisation of these factors as irrational is based on normative considerations about how behaviour ought to be and about which elements should affect agents’ preference formation. In that respect, rejecting nudges on the basis that they exploit irrationality proves to be problematic in light of the ongoing discussion and disagreement within decision-making research.

Conceptualising nudges as interventions that exploit irrationality requires accepting the rational choice model as the normative model of behaviour and distinguishing good (rational) influences from bad (non-rational) ones. However, there is considerable disagreement about identifying rational and non-rational factors and controversy about which nudges trigger rational or non-rational mechanisms. The objection to rationality already excludes nudges that most resemble rational persuasion, such as reminders, warnings, some types of information and educational campaigns. However, doubts about how nudges relate to rationality are also relevant for interventions that exploit cognitive weaknesses and are typically classified as nudges.

### *3.1.2. Rationality and autonomy*

The challenge of identifying rational and non-rational factors is underpinned by an idealistic view of decision-making and rationality. Critics identify nudges as negative influences because they do not engage rationality. Rationality appears to be the crucial element of autonomous choice within this approach; individuals are expected to engage in rational thinking, process all the relevant information and act according to their consistent preferences. However, empirical evidence on decision-making

---

<sup>1</sup>In some cases, it is not even a different concept of rationality, but just using rational choice theory, with some informational assumptions, to explain the phenomenon. For example, following the behaviour of others can be explained with the rational imitation mechanism, in which, in contexts of limited information, what others are doing signals the rational (or utility maximation) course of action (Hedström, 2006)

indicates that the approach relies on a somewhat idealistic and heroic view of rationality and decision-making, which does not match reality.

Rational choice theory is a *substantive* theory of rationality (Simon, 1985, 1986). It uses a specific set of axioms to study and model behaviour, and employs assumptions about agents' internal decision-making processes, namely perfect rationality, perfect foresight, consistent preferences and unlimited computation abilities. The expected utility model works with these axioms and explains and predicts behaviour *as if* subjects behave accordingly. Rational choice theory serves as a theory for modelling behaviour but does not account for the actual process of decision-making. It does not delve deeper into the decision-making process and does not explore the truthfulness of axioms.

While a substantive theory of rationality can be useful in decision-making research, particularly when it comes to constructing models and predicting behaviour, when the same conception is used to inform the ethical debate on autonomy, it is much more problematical. The rational choice theory model faces significant criticism both in descriptive and normative terms. Extensive experimental findings question its descriptive power (e.g. Angner & Loewenstein 2007; DellaVigna 2009), and the framework attracts significant criticism when used as a normative theory of decision-making (Elqayam & Evans 2011; Gigerenzer 2015). A conception of autonomy as rationality based on a model that does not accurately describe agents' decision-making and is notorious for its detractors is not suitable for judging the moral acceptability of nudges.

Those committed to the standard model of rationality may still argue that nudges are problematic. They may argue that, ideally, people should be rationally persuaded, and that governments and policy institutions should prioritise the use of interventions that influence behaviour via rational reasons such as education, information, monetary incentives and coercive measures. However, committing to rational choice has some limitations.

Firstly, non-rational factors will always affect behaviour. Most actions are affected by a multiplicity of motives; some of these motives qualify as rational, others not. We cannot expect people to only act for rational reasons; non-rational reasons play an equal part in and are relevant to decision-making. Even in the absence of nudges, non-rational factors usually affect choices; so, from a pragmatic standpoint, it is unclear why these factors should be considered unsuitable for use in changing decisions. Secondly, nudges do not intend to replace other strategies, and their use is compatible with trying to change people behaviour by using rational strategies. Likewise, in some cases, nudges may have some comparative advantages to other alternatives.

Recent reviews on the performance of nudges indicate that they are useful in providing behavioural change across different domains (Hummel & Maedche, 2019) and that their cost-benefit ratio can be better than those resulting from employing traditional interventions (Benartzi et al., 2017). Likewise, nudges can be more effective and less costly than other interventions.

Education campaigns have limitations when it comes to tackling some issues, are more expensive and do not always induce behavioural change in the short term (Conly, 2013; Datta & Mullainathan, 2014). In particular, the ineffectiveness of information and educational campaigns seems more evident for issues in which individuals have enough knowledge about the consequences of their choices but still fail to adapt their behaviour, for instance, in matters of health behaviour. When there is a problem of intention vs action, information and education fall short of providing sufficient change; nudges could be a more effective option (Thaler & Sunstein, 2008).

The use of more coercive strategies such as regulations, sanctions and monetary incentives tends to be more effective when it comes to changing behaviour. However, these measures are also susceptible to pitfalls that nudges may avoid. Under rational choice assumptions, in institutional design it is assumed that agents have no preferences for compliance in the absence of sanctions and incentives. However, Pettit (1996) offers a comprehensive discussion on how a deviant-centred strategy can backfire due to crowding out non-egocentric deliberation and intrinsically motivated compliance. A deterrence approach signals to agents that the situation in terms of compliance is not satisfactory for them, that the default behaviour of other agents is to avoid compliance and directs their attention to the possibility of non-compliance that they themselves would not have detected if it were not for the sanctions. In some contexts, in which this is more likely to occur, the use of nudges could avoid these adverse effects.

Similarly, problems may arise from using monetary incentives. The provision of monetary incentives provides extrinsic motivations for certain behaviours and, as such, may replace the intrinsic motivation to keep the desired behaviour once the incentives are removed. Likewise, their use in some domains might be problematic, particularly domains with moral connotations or image concerns. Monetary incentives tend to discourage those who are naturally drawn to comply, frame a situation as a monetary issue which legitimises free rider behaviour and may reduce the chances of peer punishment for non-compliance (Gneezy et al., 2011). Consequently, the use of nudges can be useful in avoiding these motivational issues.

Finally, nudges may also be beneficial when the use of monetary incentives raises concerns about matters of cost and inequality. Taxes on specific goods can have a more significant impact on people from low-socioeconomic backgrounds. For instance, research on health risk behaviour indicates that low-income individuals are more likely to purchase unhealthy products. Thus, taxing these products implies taxing the poor. Because nudges do not increase the economic cost of undesired options, their use avoids exacerbating economic inequality.

The rationality objection essentially argues that nudges do not respect rational decision-making and undermine people's autonomy. These objections only hold true under assumptions on behaviour based on the rational choice model. According to the model, individuals ought to behave and actually do behave rationally; thus, nudges that exploit irrationality compromise their autonomy. However, when

one acknowledges the limitations of rational choice theory and takes alternative understandings of rationality into consideration, the rationality objection loses strength (Schmidt, 2019). Firstly, arguing that nudges trigger irrational mechanisms is not a clear-cut empirical distinction but a normative one which requires a normative assessment of decision-making factors according to the rational choice theory framework. Secondly, the account of autonomy as rationality is too idealistic and normatively charged, and empirical research emphasises that it is based on unrealistic assumptions (Felsen & Reiner, 2015; Mills, 2015; Schubert, 2015; Schmidt, 2019).

Critics of nudging tend to point out the limitations of both rational theory and behavioural economics to explain behaviour, and strongly emphasise that the idea of Econs is an unrealistic framework for understanding human decisions. Paradoxically, at the same time, they argue against nudges due to their effect on autonomy by relying on an ideal notion of autonomy. Building on a super-agent understanding of autonomy (Schubert, 2015, p. 13), the majority of detractors of nudges defend rational persuasion and education as the ideal ways in which to influence and induce behavioural change (Gigerenzer, 2015; Hausman & Welch, 2010; White, 2013), strategies that fit in perfectly with the classical economic framework. Evidence shows that, in some cases, these strategies do not always work and struggle to induce behavioural change in the short term.

Addressing the rationality objection clears out some of the objections to nudging. However, some questions on the acceptability nudges remain unanswered. Authors who point out that the underlying mechanisms of nudge interventions are problematic in terms of rationality seem aware of the difficulties in arguing that a wide range of factors affecting decision-making should not be used to change people's choices. For instance, Hausman & Welch (2010, p. 126) argue 'Why shouldn't these factors influence preferences? Why should these factors be regarded as interferences with rational choice rather than as rational determinants of choice?'

Similarly, critics concerned about nudges taking advantage of irrational psychological mechanisms seem particularly worried about nudges that exploit mechanisms from the "biases and blunders" category and nudges that employ negative emotions. By contrast, nudges that mobilise social norms, and nudges that set commitment devices, use text warnings and use reminders raise less controversy. These ideas indicate two things: firstly, critics could be employing rationality as a proxy for conscious and reflective decision-making. Secondly, some nudges appear more problematic than others in this realm. The following section addresses these issues.

### **3.2. The reflection objection**

The reflection objection states that nudges are harmful and compromise individuals' autonomy because they actively reduce agents' engagement in the decision-making process. Reflection means that agents are aware of the factors that influence their choices, engage with these factors and exert some sort of effort and deliberation into decision-making. Some authors stress that nudges are unnoticeable, work

unconsciously and steer people towards a specific choice with limited consciousness and limited effort. As a result, critics argue that nudges compromise people's reflectiveness.

The worries on reflection respond to the idea that for people to enjoy self-determination and control and manage their own decisions, they need to engage their reflective and reasoning capacities in decision-making. In the absence of reflectiveness, people no longer make their own choices, and it is easy for policymakers to manipulate them into doing things they would not do if the influence were open and fully engaging with their reflective capacities. In this case, the focus is not on nudges compromising rationality and exploiting irrational biases; instead, it is on reflective decision-making. For a nudge to be ethically permissible, it has to engage some form of conscious deliberation. Influence has to be perceived, compared and valued against other reasons. Non-ethically permissible influence is influence that overrides consciousness, is difficult to perceive and minimizes agents' roles and efforts in the decision-making process.

Two features of nudges frequently come up when discussing issues of reflectiveness: how nudges engage deliberation and whether nudges are transparent. Objections about both features claim that nudges compromise individual autonomy when they influence behaviour by bypassing people's capacity for deliberation and do so in a way that is not easily recognisable or disclosed to the subjects (Conly, 2013; White, 2013).

Blumenthal-Barby and Burroughs (2012), state that nudges can imply manipulation when they are difficult to perceive:

‘Manipulation occurs when one influences another by bypassing their capacity for reason, either by exploiting nonrational elements of psychological makeup or by influencing choices in a way that is not obvious to the subject’ (p. 5)

Bovens (2009) notes that nudges tend to be hard to perceived and work best when unnoticeable:

‘The psychological mechanisms that are exploited in Cafeteria and in Save More tomorrow<sup>2</sup> typically work better in the dark. If we tell students that the order of the food in the Cafeteria is rearranged for dietary purposes, then the intervention may be less successful’ (p. 3)

Grüne-Yanoff and Hertwig (2016) state that nudges can produce an automatic response without agents' active attention:

‘in a nudge, the policymaker does not rely on the agent's ability to stop or override a targeted behavior or cognition. Instead, the nudge intervention can ‘remedy’ an individual's actions

---

<sup>2</sup> The cafeteria nudge references the reorganization of the food chain in a Cafeteria in a way that makes healthy options easier and more convenient to select. The Save More tomorrow references a programme that increases pension contributions with each pay raise by default (Thaler & Sunstein, 2008)

without the individual making an active contribution. It is this feature that leads critics to argue that nudges are manipulative, and that they violate autonomy and dignity' (p. 176)

Baldwin (2014) states that some nudges compromise autonomy because they work without awareness and reflection:

'the targeted individual's behavioural or volitional limitations and 'automatic' responses will in practice lead him or her to accept the nudge with limited awareness and reflection.' (p. 836)

These quotes express worries about nudges impeding reflectiveness. In contrast to the examples discussed in the last section, it is easy to appreciate the difference with worries on rationality. For instance, nudging by using framing effects is an intrusion on autonomy because agents cannot identify that there is an intervention aimed at shaping their behaviour and fail to recognise how the frame affected their decision. Baldwin (2014) argues that framing devices can shape behaviour 'in a manner that is resistant to unpacking in so far as assessing the nature and extent of the nudge is not readily achieved by reflection' (p. 836). In this case, the worry is not about rationality but about reflectiveness. Similarly, defaults appear problematic because they can work without individuals perceiving the intervention or its effect. Therefore, defaults undermine autonomy, not because they exploit a cognitive bias, but due to their potential to influence choices without awareness and reflection (Smith, Goldstein, and Johnson 2013). By contrast, the communication of social norms is a nudge that does not appear as problematic because it engages reflection: agents can identify the nudge, and it works through a process that involves some degree of reflection (Bovens, 2009; Hansen & Jespersen, 2013). As a result, even though communicating social norms might activate a non-rational response, it is a less problematic nudge because it engages reflectiveness.

The worries on reflection underpin worries on manipulation, transparency, and dignity. I am not going to address these objections specifically because I suggest that they boil down to worries about nudges' effect on reflectiveness. For instance, the account of manipulation that is used in nudge literature always references nudges working unconsciously and 'in the dark', i.e., working without the full attention and awareness of agents. Because the influence operates in the dark, the nudge is non-transparent and implies manipulation.

However, manipulation can be understood in different ways and can take different forms. Firstly, the concept of manipulation implies that A has to pressure B to do something that B does not want to do. In the case of nudges, while we cannot argue that people always want to be nudged towards an outcome, we cannot say that they oppose it. Likewise, manipulation can happen transparently, by shaping arguments, exploiting emotions and deceiving people. In these forms, manipulation does not operate 'in the dark' and has nothing to do with whether agents notice the effect or influence of an intervention (Mitchell, 2004). In nudge literature, claims of manipulation and transparency relate to the degree of

reflectiveness that agents have in decision-making; thus, it is better to address these objections by examining the effects of nudges on reflectiveness.

Similarly, objections related to dignity and individual responsibility also reflect concerns about how nudges affect reflectiveness. Detractors of nudges maintain that nudges do not help us to learn to make better choices in the future, because they override our conscious reasoning and block the learning process (White, 2013, p. 102). This leads some to worry that nudging infantilises individuals and undermines their responsibility and dignity (Bovens, 2009; Gigerenzer, 2015; White, 2013). Again, these objections raise concerns about how nudges shape choices, and how much reflection they engage; therefore, to allay such worries, it would be better to address them by discussing the effects of nudges on reflectiveness.

The key aspects of concern regarding the reflection objection can be listed as follows: (a) nudges bypass deliberation and affect behaviour unconsciously, and (b) nudges lack transparency and have to operate unnoticeably to work. These two features raise concerns about nudges pushing people into doing things without reflectiveness, thereby undermining their autonomy. The reflection objection has two main problems, just as the rationality objection does. Firstly, it is unclear which nudges bypass deliberation, and are unnoticeable. Secondly, it considers reflectiveness to be a crucial component of autonomous choice, a notion which appears to be too idealistic when the psychological evidence is taken into account. I shall start by commenting on the first issue, and then return to the discussion on reflectiveness, decision-making and autonomy.

### *3.2.1. Diversity of nudges and ethical considerations.*

With regard to nudges impeding people's reflectiveness, there appears to be a consensus on the features that make nudges problematic: (a) they bypass deliberation (b) they are not transparent. While both claims are frequent in nudge literature, many authors also acknowledge that not all nudges have these features. Sunstein points this out by emphasising 'the importance of having a sufficiently capacious sense of the category of nudges, and an appreciation of the differences among them' (Sunstein 2015a, p. 513) and uses this argument to defend nudges against objections. In line with his argument, many authors have paid more attention to the fact that conclusions about the moral acceptability of nudges should focus more on the specific features of each nudge.

Current literature offers different classifications of nudge interventions. In general, the classifications take into account how nudges involve deliberation, and whether nudges are transparent as standard features for assessing their ethical acceptability. In all cases, the implicit goal is to find good and bad nudges; i.e. nudges which, due to their characteristics, are morally acceptable, and nudges that fail to respect reflection, i.e., fail to respect autonomy. Therefore, looking at the characteristics of nudges seems to be relevant when judging their ethical acceptability. Currently, there are diverse classifications. These classifications share intuitions and ideas; however, they also seem to use different

criteria and tend to provide nonmatching classifications. The sections that follow discuss some of the classifications and their limitations in terms of the ethical debate.

- *Distinction between System 1 and System 2*

The most well-received and commonly used classification of nudges is the differentiation between System 1 and System 2 nudges. In the discussion on the ethical implications of nudges, there is a widespread intuition that distinguishes between two types of nudges. Some nudges operate in more conscious and reflective ways, need to be transparent and require effort from the nudgees to work. By contrast, some nudges work automatically, unconsciously, covertly and do not require effort from nudgees to change behaviour effectively. The distinction can be found in research on the empirical performance and effectiveness of nudge interventions (e.g. Hollands et al. 2013; Smith, Goldstein and Johnson 2013) and is implicitly present in the theoretical framework underlying nudging (e.g., Kahneman 2011). The distinction between System 1 and System 2 nudges includes these ideas.

The distinction between System 1 and System 2 follows Kahneman's (2011) account of the dual-process cognitive theory, which describes two distinct systems underlying human reasoning: System 1 and System 2. System 1 is fast, automatic, uncontrolled, and unconscious. System 2 is slow, conscious, reflective, and controlled (Kahneman 2011). The approach perfectly fits with the debate and characterisation of nudges, since Kahneman's (2011) account of the dual-process cognitive theory is an inherent part of the original characterisation of nudges (Heilmann 2014; Thaler and Sunstein 2008). When applied to nudges, the schema distinguishes between System 1 and System 2 nudges. System 1 nudges tend to be described as unconscious, non-deliberative influences that override human agency. System 2 nudges are described as reflective triggering influences that engage agency. The two types are also referred to as non-educational and educational nudges respectively (e.g. Hertwig and Ryall 2016; Sunstein 2017c). Examples of System 1 nudges include the use of defaults and the design of physical options in the decision-making environment. Examples of System 2 nudges include disclosing information or using educational campaigns. System 2 nudges implicitly tend to be seen as good nudges while System 1 nudges appear to be more morally problematic.

The distinction between System 1 and System 2 serves as a good schema to argue against objections that present nudges as interventions which always bypass deliberation, tend to prompt an unconscious or unreflective response and are usually unnoticed. The distinction has been adopted by proponents of nudges and is regularly employed to differentiate between nudge interventions and address their ethical implications. For instance, the distinction between System 1 and System 2 is often used in previous literature on *attitudes towards nudges* to study whether people accept nudges and how they judge different nudge interventions (e.g. Felsen, Castelo, and Reiner 2013; Hagman et al. 2015; Jung and Mellers 2016; Sunstein 2017c). These studies describe System 1 nudges as covert, subconscious and automatic and System 2 nudges as overt, conscious-driven influences to choose, 'educational

campaigns' or interventions that 'provide information' to make better-informed choices (Hertwig, 2017; Sunstein, 2017c). In this literature, the distinction is not exhaustive, but it is consistent. The authors seem to agree on how to describe System 1 and System 2 nudges, which interventions fit each label, and why this is considered a relevant differentiation for the ethical debate. These elements strengthen the argument in favour of using the distinction, and it is, indeed, being used more. However, it is being used without explicit criteria, without references to what underpins the classification of a nudge as a System 1 or System 2 intervention and without evidence about how the different nudges engage the cognitive mechanisms of each system. Therefore, despite this broad adoption, there are significant drawbacks.

Firstly, dual-process theories of human reasoning pose a few theoretical challenges. While such approaches may be compelling, they often include different understandings and approximations of what System 1 and System 2 cognition entails (Evans 2008, 2003). Notably, the consideration that System 1 (unconscious) is generally incorrect, and that System 2 (conscious) is generally correct is highly contested. Evidence suggests that heuristics that operate unconsciously might produce good decisions and that they can also work consciously. Likewise, Melnikoff and Bargh (2018) stress that dual-process theories lack empirical evidence to support the idea that attributes typically associated with System 1 (e.g., unconscious, unintentional, and incorrect) and System 2 (e.g., conscious, reflective, and correct) align. My intention here is not to review the limitations of dual-system theories or to provide a classification of nudges according to the distinction. Instead, I am questioning the potential of the System 1 - System 2 schema to serve as a basis for drawing normative conclusions about nudges.

Secondly, theoretical inconsistency poses problems when it comes to sorting out nudges and classifying them as System 1 or System 2 interventions. In previous literature, there are no empirically grounded attempts that describe and classify nudges as System 1 or System 2 (Mongin and Cozic 2018). Heilmann (2014) acknowledges that, in order to withstand methodological and epistemic criticism about nudges, one needs to fully endorse dual-process cognitive theories, yet he does not discuss how the different nudges engage with System 1 and System 2 mechanisms. Likewise, Sunstein acknowledges that some nudges relate to System 1 and other interventions seem to apply more to System 2. However, he also states that nudges might trigger or exploit System 1, work due to its effect, or be designed to prevent its mistakes. In his work, the relation between nudges and systems is diffuse and inconclusive, and he goes on to emphasise that the distinction lacks psychological foundations (Sunstein 2015b, 2016). Sunstein's examples of System 2 nudges (or educational nudges) resemble the examples that he uses to defend nudges against the rationality objection. The nudges that respect rationality are the same nudges that engage System 2; (educational campaigns, reminders, the provision of feedback and the use of warnings). The System 1 category includes interventions such as defaults and the reorganization of choice architecture, together with examples that tend to be seen as nudges that compromise rationality. This characterisation leaves several nudges in undefined territory, for instance, interventions that

communicate social norms. The theoretical debate makes it hard to classify nudges between System 1 and System 2 and proves to be as challenging as identifying failures of rationality.

In practice, the distinction between System 1 and System 2 lacks robustness and tends to reflect intuitions and assumptions about the mental processes at work when nudging. The lack of actual psychological foundations to sorting nudges between System 1 and System 2 indicates that the distinction between System 1 and System 2 may be a useful label for distinguishing between good and bad nudges rather than as a proper empirically founded classification of nudges. The System 1 label usually references interventions that trigger worries of non-transparency, manipulation and unconsciousness. By contrast, the System 2 label is used for referring to interventions that intuitively appear to be more transparent, engage reflection and less problematic overall. While the differentiation might be intuitive and attractive, it is challenging to use for categorising nudges and has limitations when it comes to working as a reliable classification to address the ethical debate.

- *Baldwin's taxonomy of nudges in three degrees*

Baldwin (2014) develops a taxonomy of nudge interventions that classifies nudges in three degrees: First Degree, Second Degree and Third Degree nudges, depending on how an intervention engages autonomy. Baldwin operationalises autonomy as the way in which nudges take reflective decision-making into account (Baldwin 2014, p.835). First Degree nudges are interventions that engage and encourage reflective thinking, for instance, reminders and the provision of feedback. Baldwin argues that these interventions promote informed choices and are unproblematic in terms of autonomy. Second Degree nudges are interventions that build on behavioural limitations to shape decisions. As examples, Baldwin considers default rules and the reorganisation of choice architecture (Baldwin 2014, p. 835). Baldwin argues that Second Degree nudges are more problematic for autonomy because they seek to exploit an 'automatic response' and intend to promote choices made 'with limited awareness and reflection' (Baldwin 2014, p. 836). However, Second Degree nudges are somehow unproblematic because, with some effort, and attention, agents can unpack the nudge and resist its influence. Finally, Baldwin defines Third Degree nudges as interventions that inhibit or obstruct reflection. Third Degree nudges work in ways whereby agents cannot identify to what extent the nudge has shaped their preferences and represent a severe intrusion on autonomy. As examples of a Third Degree nudge, Baldwin includes a campaign to encourage healthy eating with the message "Don't lose your looks, junk the junk food!" and the use of subliminal advertising. Baldwin argues that in both cases, the cognitive effect of the interventions is so strong that agents cannot identify to what extent their preferences are affected by the nudge. The "Don't lose your looks, junk the junk food!" campaign exploits loss aversions. Baldwin contends that, even though this campaign expects some reflection from the nudgees, i.e., it expects them to choose healthier options, the message creates such a strong negative emotion that agents fail to unpack how much the nudge affected their preferences. Similarly, subliminal

advertising qualifies as a Third Degree nudge because agents also fail to discern the reason behind their change in preferences.

Baldwin's taxonomy considers the two main themes within the reflection objection. First Degree nudges involve deliberation and are usually perceived; thus, they respect people's reflective capacities. Second Degree nudges can potentially bypass deliberation and work unnoticed; therefore, they work with limited reflection. Finally, Third Degree nudges obstruct reflection and can bypass deliberation and work unnoticed; thus, they are very much more problematic in terms of reflectiveness and autonomy.

- *Saghai's amendment concept of nudges*

In an alternative approach, Saghai (2013) redefines the concept of a nudge by taking into account how much an influence respects agents' freedom of choice. Saghai draws a distinction between nudges and *behavioural prods*. In his approach, a nudge is a 'substantially non-controlling' influence. By contrast, a behavioural prod is an influence that uses the same mechanisms, but it is substantially controlling and fails to respect agents' freedom of choice.

Saghai establishes three conditions to identify substantially non-controlling influences: the influence must be I) easily resistible and has to II) respect attention-bringing capacities and inhibitory capacities. He defines these conditions in the following terms:

- I. A's influence is easily resistible if B is able to effortlessly oppose the pressure to get them to w if they do not want to go to w.
- II. B is easily able to resist A's influence when: i) B has the capacity to become aware of A's pressure to get them to w (attention-bringing capacities), and ii) B has the capacity to inhibit their triggered propensity to w (inhibitory capacities).

According to Saghai, interventions such as the "cafeteria nudge", the prescription of generic medication by default and the "Less Than You Think" campaign<sup>3</sup> are nudges. The "cafeteria nudge" and the prescription of generic medication by default are substantially non-controlling influences because they respect attention-bringing capacities and agents can easily oppose the pressure exerted by the nudge. According to Saghai, in both cases, agents can easily choose other food and medication options. The "Less Than You Think" campaign is also a nudge because agents are aware of the pressure and intention of nudge (I) (a campaign to reduce alcohol consumption) and (II) to inhibit the trigger propensity (drink less or more).

---

<sup>3</sup> The "Less Than You Think" campaign is an informative campaign that states the percentage of binge drinking among students at the university campuses in Montana during weekdays. The campaign illustrates that the majority of students binge drink less often than what most assume. The message breaks a case of pluralistic ignorance and uses social norms to reduce alcohol consumption.

Alternatively, Saghai argues that the ‘Less Than You Think’ campaign but using false information and the triggering of sensorial responses that produce a positive emotional reaction, are behavioural prods. In the first case, the information is false, and the nudger is deceiving the nudgees; in the second case, the intervention is concealed, and agents cannot inhibit its triggered propensity. In both cases, agents cannot easily resist the interventions, and the interventions are not nudges.

Saghai’s approach also contains critical elements within the reflectiveness objection: attention-bringing capacities refer to whether agents perceived the nudge (transparency); inhibitory capacities refer to whether nudgees can resist its cognitive pressure (effect on deliberation).

- *Hansen and Jespersen’s epistemic distinction of nudges*

Hansen & Jespersen (2013) present an epistemic distinction of nudges according to how different nudges engage rational deliberation and whether the interventions are transparent. With regard to deliberation, they classify nudges as Type 1 and Type 2. The aim of Type 1 nudges is to change behaviour without involving reflective thinking, while Type 2 nudges involve reflective thinking, and mainly influence behaviour resulting from some degree of deliberation. Examples of Type 1 nudges include the reorganisation of a consumer environment to make (un)healthy food more (in)visible or (in)convenient to find, the alteration of the size of the food recipient and the automatic default rules that prompt people to sign up as organ donors. Some examples of Type 2 nudges under the authors’ typology include the ‘look right’ warning in the streets of London, the ‘fly in the urinal’ and energy bills which use a social comparison of consumption. Type 1 nudges are aimed at influencing behaviour caused by automatic thinking, whereas Type 2 nudges affect the attention and premises of choice, thus involving reflective thinking (Hansen & Jespersen, 2013; Hollands et al., 2013). Type 1 nudges are considered more ethically problematic, as they may constitute manipulation of choice.

Hansen and Jespersen (2013) add considerations about the degree of transparency of nudges to the first dimension. In their framework, a transparent nudge is a ‘nudge provided in such a way that the intention behind it, as well as the means by which behavioural change is pursued, could reasonably be expected to be transparent to the agent being nudged’ (Hansen & Jespersen 2013, p. 17). They successfully argue that interventions on choice architecture are noticeable, regardless of their efficacy. Interventions such as warnings signs, reminders, and social comparison, are noticeable and work precisely because of this. These interventions cannot work if subjects are unable to perceive them and are easily recognisable as behavioural change interventions<sup>4</sup>. In their classification, both Type 1 and Type 2, nudges can be

---

<sup>4</sup> In Hansen & Jespersen (2013), ‘announcing “on time” when trains arrive on time’ and ‘the playing of relaxing music while passengers board a plane in order to calm them’ count as transparent nudges. I believe that rather than transparent, these types of nudges are perceptible as people can notice the music or hear the ‘on-time’ announcement; however, it is less clear whether they will perceive the effect such interventions have on their behaviour. For the means to be appropriately transparent, I suggest that not only the presence of nudges but also their nature as a behavioural change input ought to be evident.

transparent and un-transparent. Announcing “on time” when trains arrive on time’ and ‘the playing of relaxing music while passengers board a plane to calm them’ count as transparent nudges. In these situations, subjects can easily identify the nudge and its intention. By contrast, the framing of information in ways that exploit loss aversion and the use of reference numbers as anchors to influence choices are non-transparent nudges, since subjects cannot identify the intervention or the effect it has on their behaviour. The authors argue that non-transparent nudges are morally retrievable and more likely to manipulate choices. If a nudge intervention is easily perceived as an attempt to shape behaviour, it is less problematic. By contrast, some nudge interventions can go completely unnoticed by the subjects and still lead them towards a particular outcome<sup>5</sup>. As mentioned, these cases raise more ethical doubts. Again, under their classification, deliberation and transparency are crucial elements when it comes to discussing the potential of nudges to operate with or without awareness and reflection. Under their classification, a Type 1 non-transparent nudge is deemed to be potentially manipulating choices.

- *Bovens (2009) A criterion of nudges transparency*

Bovens (2009) explores the differences between nudges in terms of transparency. In the original characterisation of nudges, Thaler and Sunstein (2008) acknowledge that transparency is a necessity to ensure that nudges are used to do good. In terms of transparency, they defend nudges that rely on Rawls’s publicity principle, which establishes that a government should not adopt policies that it is not ‘willing or able to defend publicly to its own citizens’ (Thaler & Sunstein, 2008, p. 244).

Nevertheless, Thaler and Sunstein’s stance on transparency does not address the concern that subjects often fail to recognise the presence of an intervention. The discussion on transparency implies two different arguments: whether a nudge is or can be subject to accountability and public scrutiny, and whether an intervention is or can be identified by the subjects that are exposed to it. Bovens (2009) distinguishes between *interference transparency* (governments being transparent about the means that they employ to change behaviour) and *token interference transparency* (the likelihood that nudges will be able to identify the nudge and its desired outcome). Building on this distinction, he suggests that interference transparency is feasible but not sufficient to ensure that nudges will not be employed for unethical results; manipulation is still possible if nudges do not satisfy token interference transparency. The fact that nudges are often not directly evident to nudges is taken as a feature with negative ethical implication.

Building on the distinction between interference transparency and token interference transparency, Bovens (2009) argues that nudges vary in terms of ethical acceptability. Nudges that are not directly evident to the subjects are more morally problematic. Nudges that are noticeable in the situation in

---

<sup>5</sup> Please note that not being aware of the presence of an intervention is not the same as not being aware of the effect or result of an intervention.

which they work are more acceptable. Bovens argues that social advertising that provides information or affects emotions is a nudge that satisfies token interference transparency. In this case, the aims employed to achieve behavioural change are apparent to agents, meaning that agents are able to recognise that pressure is being exerted to affect their choices, and they can usually understand the intention behind it. By contrast, Bovens argues that the “cafeteria nudge” lacks token transparency because, in the cafeteria, subjects cannot identify the intervention or its intended effect. According to Bovens, the token interference transparency of nudges cannot be improved, because ‘when the existence and purpose of a choice architecture intervention are disclosed, the intervention may be less successful’ (Bovens, 2009, p. 217). Bovens’ distinction between interference transparency and token interference transparency also emphasises the importance of nudges retaining awareness over the factors that affect their choices.

The classifications described above rely on the same underlying idea: nudges vary in form and, therefore, they also do in terms of ethical acceptability. The five classifications share similar intuitions about the features that make a nudge more acceptable or less so: how nudges involve deliberation and whether they are easy to identify. However, the classifications differ significantly; authors operationalise the criteria in different ways and present different interpretations of the moral value of the same interventions. Table 3 summarises the classifications and serves to display the disparity of criteria and categories and the different interpretations of the same nudge intervention as a good or bad nudge.

**Table 3. Summary of classifications of nudges and their ethical implications.**

Classifications	Categories	Criteria	Ethical acceptability	Examples
System 1 - System 2 nudges	System 1 nudges	Engages System 1. Can work unconsciously, with limited awareness and reflection.	<b>Somewhat problematic.</b> Undermines people autonomy	Defaults, an opt-out organ donor regime, re-organisations of the physical environment
	System 2 nudges	Engages System 2. Involves reflection.	<b>Unproblematic</b>	Informational and educational campaigns, warnings, reminders, feedback.
Baldwin (2014)  Three Degree nudge taxonomy	First Degree nudge	Involves deliberative capacities and engages and encourages reflective thinking	<b>Unproblematic.</b> Respects autonomy and decision-making.	Reminders, provision of simple information, the provision of feedback, text health warnings on cigarette packaging.
	Second Degree nudge	Exploits behavioural limitations to shape behaviour	<b>Somewhat problematic.</b> Negative effect on individual autonomy. Seeks an automatic response. Agents can accept it with limited awareness and reflection, but on reflection, agents can perceive the nudge and counteract its effect.	Defaults, an opt-out organ donor regime, re-organisations of the physical environment.
	Third Degree nudge	Obstructs reflections, seeks negative emotional responses, undisclosed and difficult to perceive influences.	<b>Very problematic.</b> Agents fail to unpack the nature and extent of the decision or preference shaping.	Shocking images on cigarette packaging, unpublicised subliminal TV messages. An advertising campaign that promotes healthy eating with the slogan: 'Don't lose your looks, junk the junk food!.'
Saghai (2013) An amendment concept of nudges	Nudges	Substantially non-controlling influences. Easily resistible influence, that agents can (i) become aware of (attention-bringing capacities) and (ii) can inhibit the propensity that they trigger (inhibitory capacities)	<b>Unproblematic.</b> Influences that robustly preserve freedom of choice.	The cafeteria nudge; the prescription of generic medication by default; the 'Less Than You Think' campaign on binge drinking.
	Behavioural prods	Substantially controlling. Influences that do not respect (i) attention-bringing capacities and (ii) inhibitory capacities	<b>Problematic.</b> Influences that do not respect freedom of choice.	The 'LTYT' campaign but using false information, a pharmaceutical TV ad that lists the side effects of the medication in a monotonous voice while displaying images of butterflies and happy people to counteract the adverse effects.

Classifications	Categories	Criteria	Ethical acceptability	Examples
Hansen & Jespersen (2013)  An epistemic distinction of nudges	Type 1 transparent nudges	(i) They change behaviour without involving reflective thinking and (ii) Their intention and mechanisms can be perceived by those nudged	<b>Somewhat problematic.</b> They do not engage reflective thinking. Behavioural change is automatic Transparent influence that involves manipulation of behaviour.	Playing relaxing music in plains or train stations to avoid stress, explicit optical illusions to control traffic, changing printer defaults, announcing “on time” in trains.
	Type 1 non-transparent nudges	(i) They change behaviour without involving reflective thinking and (ii) Their intention and mechanisms are non-transparent to those nudged	<b>Problematic.</b> Behavioural change without involving reflective thinking which is unlikely to be recognised. Non-transparent manipulation of behaviour	Opt-out system for organ donation, altering product sizes to reduce consumption, implicit optical illusions to control traffic, automatic enrolment in programmes, and organisation of the physical environment.
	Type 2 transparent nudges	(i) They change behaviour involving reflective thinking (ii) Their intention and means are transparent to those nudged (transparent)	<b>Unproblematic.</b> They engage the reflective system and make it easy for subjects to reconstruct the nudge’s intentions and mechanisms. Facilitate consistent choices.	Text warnings, such as “look right” in London, labelling of calories, energy bills using social comparison of consumption, the fly-in-the-urinal nudge.
	Type 2 non-transparent nudges	(i) They change behaviour involving reflective thinking and influence (ii) Their intention and mechanisms are non-transparent to those nudged (transparent)	<b>Somewhat problematic.</b> They engage the reflective system but the influence and mechanisms of the nudge remain unnoticed. They involve manipulation of choice.	Framing information, adding irrelevant alternatives, arranging incentives such as a lottery, using a poster with faces to increase compliance rates, and anchoring a random number to change behaviour.
Bovens (2009)  A criterion of nudge transparency	Token interference transparency	The nudger is transparent about the means that they will employ, and the nudge is noticeable in the choice situation in which it is applied.	<b>Unproblematic.</b> Transparent nudge. Easy to become aware of the intervention and its intention.	Social advertising.
	Interference transparency	The nudger is transparent about the means that they will employ, but the nudge is not noticeable in the choice situation in which it is applied.	<b>Problematic.</b> Agents cannot recognise the nudge and its intention. Transparency cannot be improved because it reduces the effectiveness	Defaults, framing, anchoring.

Comparing the criteria used by the different classifications poses problems and identifies inconsistencies in some of the classifications. For instance, the communication of social norms is usually perceived as a non-problematic nudge. For Hansen & Jespersen (2013) the communication of social norms is a Type 2 transparent nudge. The authors contend that it is an intervention that involves reflective thinking because it affects choices through deliberation. Likewise, they classify the intervention as a transparent nudge because agents can recognise the intervention and its intention. Similarly, Bovens (2009, 2013) argues that even if the communication of social norms exploits some pattern of irrationality, the intervention satisfies token interference transparency because agents can easily recognise the nudge and its intention. Thus, the nudge is perceived as unproblematic. By contrast, Baldwin maintains that interventions that use framing devices are problematic for individual autonomy. In his account, he does not explicitly discuss the example of social norms. However, if we take his criteria into consideration, the use of framing devices to communicate social norms could pass as a Third Degree nudge. Similarly, Saghai deems the communication of social norms as a nudge if the message uses real information, whereas he finds it problematic if the messages use false information.

The “cafeteria nudge” also has a different moral profile depending on the classification. In general, it appears more problematic. Hansen & Jespersen (2013) indicate that it is a Type 1 nudge because it does not imply reflection and directly affects choices. Likewise, they classify it as a non-transparent nudges because individuals cannot perceive the means by which their behaviour is influenced. By both features, it is considered a nudge that can imply manipulation. Bovens (2009) also argues that the ‘cafeteria nudge’ lacks token interference transparency; and argues that it is an intervention that takes away individuals’ control over their choices. Baldwin (2014) classifies the ‘cafeteria nudge’ as a Second Degree nudges and argues that it is an intervention that exploits biases and expects an automatic response; thus, potentially problematic for autonomy. Contrary, Saghai (2013) argues that the nudge is substantially non-controlling because agents that dislike its aim, this is, those that would like to eat something different from what the physical distribution favours, can easily opt-out of the nudge and choose as they please. Again, as it is the case for the communication of social norms, there is disagreement over whether the ‘cafeteria nudge’ is an acceptable or unacceptable intervention.

The classification also shows disagreements with the criteria they employ to classify interventions. All of the classifications discussed above judge nudges moral acceptability in terms of how nudges operate; however, they vary in the type of criteria they used. To distinguish nudges, authors tend to focus on two types of criteria: cognitive factors or design factors. Those that focus on cognitive factors intend to differentiate nudges and value their ethical acceptability by looking at the cognitive mechanisms that a nudge trigger to work. Those that focus on design elements look at the characteristics of the interventions; this is, features about their design. The differentiation of criteria is independent of how well the different authors fundament their distinctions; it solely refers to the elements that are taken as central to classify nudges.

Considering the previous classifications; the System 1 vs System 2 distinction is a cognitive-based distinction. It tries to differentiate nudges and establish their moral status according to the underlying cognitive process that makes a nudge work. Contrary, Bovens (2009), and Hansen and Jespersen (2013) approaches judge nudges moral acceptability by taking into account design criteria. For Bovens (2009), some nudges satisfy token interference transparency because agents can quickly grasp the presence of an intervention and its intention. According to Bovens, whether a nudge satisfies the criteria of token interference transparency is independent on the cognitive mechanism that it triggers. For instance, he argues that, while social advertising exploits irrational psychological mechanisms the fact that it is easy for the subjects to identify the intervention and reconstruct its intention, implies that the nudge satisfies token interference transparency and it is not as harmful as other nudges.

Similarly, Hansen and Jespersen (2013) try to differentiate nudges according to substantive criteria. In their classification, they explicitly state that the distinction between Type 1 and Type 2 nudges is not based on cognitive mechanism but on whether nudges intend to involve or avoid reflective thinking. Likewise, they present an epistemic transparency criterion by which they argue that nudges are transparent if subjects can identify the intervention and its intention. In this case, their interpretation of the two criteria (deliberation and transparency) does not involve considerations about cognitive mechanisms.

**Table 4. Types of classifications of nudges according to their type of criteria.**

<b>Distinctions</b>	<b>Description</b>	<b>Cases</b>
Cognitive distinction	Classifies nudges according to the underlying cognitive process	The System 1 – System 2 distinction Baldwin’s Three Degrees of nudges Saghai’s new concept of nudge
Design distinction	Classifies nudges according to features of the intervention	Hansen & Jespersen’s (2013) epistemic distinction of nudges Bovens’ <i>token interference transparency</i> criterion

Some classifications, however, pose problems with the criteria they employ. The classifications used by Baldwin (2014) and Saghai (2013) are good illustrations of this point. Both try to classify nudges using cognitive mechanisms but fail to provide criteria in this sense. At the same time, when evaluated according to design criteria, the interventions they deemed problematic appear unproblematic.

Starting with Baldwin’s taxonomy, as previously explained, Baldwin differentiates between First Degree, Second Degree and Third Degree nudges. A priori, his distinction seems to be based on cognitive factors; however, it lacks psychological foundations. The distinction he draws between First Degree and Second Degree nudges is based on whether nudge interventions enhance reflection and are noticed (First Degree nudges) or exploit cognitive biases and work with limited awareness (Second Degree nudges). While he references cognitive aspects, Baldwin does not go deeper into the principles

that underlie his classification; nor does he provide a classification based on psychological mechanisms. In practice, the taxonomy resembles the distinctions Hansen and Jespersen (2013) make between Type 1 and Type 2 nudges and transparent and non-transparent nudges. However, in Baldwin's case, both features are included in one category. Baldwin's First Degree nudges enhance reflection and subjects are aware of the nudge, while Second Degree nudges tend to bypass reflection and are non-transparent. Baldwin's criteria also appear problematic if we look into the Third Degree nudge category. Baldwin defines Third Degree nudges as influences that obstruct reflection in ways that make agents fail to discern how much influences affect their change in preferences. As examples of Third Degree nudges, Baldwin discusses the "Don't lose your looks, junk the junk food!" campaign, the use of shocking images of illness to reduce smoking, and subliminal TV messages. His criteria work for subliminal advertising. Subliminal advertising is an intervention that works beyond the borders of conscious perception. Yet, subliminal advertising is not a nudge. In the other examples, the criteria are not robust enough.

To illustrate this point, according to Baldwin, the "Don't lose your looks, junk the junk food!" campaign and the use of shocking images on cigarette packaging trigger strong negative emotions that make agents unable to resist them. In these cases, Baldwin states that, due to the underlying cognitive mechanism (i.e., the triggering of negative emotions), these interventions are Third Degree nudges, and very problematic in terms of autonomy.

While I can understand some of the concerns about using these interventions, I do not see how they are as worrying for individual autonomy as Baldwin states. Both the food campaign and shocking images can make nudges feel very uncomfortable. However, I am not that convinced that that obstructs deliberation and produces an unreflective change in preferences. Evidence indicates that these particular interventions are often ineffective, which supports the idea that their supposed interference in individual autonomy is not as damaging as it is presented as being. It would be an assumption to argue that a change in preference results in the unreflective effect of negative emotions.

There also seems to be a double standard on how to judge the effect of emotions on behaviour. What about interventions that trigger a strong positive emotion such as empathy or hope? Would those interventions be Third Degree nudges? Is the cognitive process different? Is their effect stronger or weaker? Cognitive research seems to suggest that negative emotions are more potent motivators than positive emotions, but these are very much unresolved questions. Similarly, what about an intervention that triggers a strong negative emotion but seems to produce a reflective long-term behavioural change?<sup>6</sup>. Research on the role of emotions and their relationship with reflection and rationality is still

---

<sup>6</sup>Bicchieri (2016) argues that instead of rationality, eliciting strong negative emotions such as fear or disgust can be a successful strategy for inducing behavioural change. For instance, to reduce defecating in public, evoking disgust causes behavioural change. While in this example we see the strong motivational effect of negative emotions, we do not see this as an unreflective change of preferences.

too scarce to support the idea that negative emotions produce an automatic and unreflective change of preferences.

There is a second critical problem regarding Baldwin's taxonomy; if we classify his examples according to design factors, the interventions that he categorises as Third Degree nudges seem to be unproblematic, contrary to his theory. For instance, taking Bovens' (2009, 2013) token interference transparency criterion, we could argue that both the "Don't lose your looks, junk the junk food!" campaign and the shocking images on cigarette packaging are interventions that satisfy token interference transparency. Agents can easily identify the interventions and grasp their intentions; therefore, these nudges are unproblematic. Similarly, taking Hansen & Jespersen's (2013) criteria, we could argue that both nudges are Type 2 transparent nudges. The interventions are transparent because agents can easily recognise the nudges and their intention, and both nudges work only if noticed (transparent nudges). Likewise, these nudges are Type 2 nudges because they intend to affect the premises of choice not to affect the choice directly. So, even if they trigger System 1 cognitive mechanisms, they need to involve some degree of reflective thinking.

In conclusion, while Baldwin seems to argue in favour of distinguishing nudges according to their underlying cognitive mechanisms, his taxonomy is not clear as to how to do so. This is notably lacking in the definition of Third Degree nudges. At the same time, when valued according to design criteria, his examples of Third Degree nudges appear to be far less disturbing in terms of autonomy than he argues.

Saghai's classification also poses a few problems. Saghai's definition of nudges states that nudges ought to be easily resistible and respect attention-bringing capacities and inhibitory capacities; however, his criteria to define both aspects are unclear. Regarding attention-bringing capacities, Saghai argues that they are activated if there is sufficient conflict between the nudgee's aims and those of the nudger. However, this condition does not detail which nudges are more or less likely to be recognised by the subjects. Nor does it detail how people who do not have pre-defined preferences for a particular aim might react to nudges. Nudges often affect decisions for which people do not have pre-defined preferences, so it is relevant to clarify how people will react to the nudge in these cases. Finally, whether those that despise the aims of nudges will effectively notice a nudge is not a fact, but an empirical question for which Saghai fails to provide a proper answer in his paper.

Saghai's criteria for inhibitory capacities are also vague. The author defines inhibitory capacities as the 'capacity to inhibit our propensity to do what the influencer wants us to do', and further indicates 'the capacity to stop a cognitive process once it is triggered' (Saghai, 2013, p. 4). By this definition, it looks as though he has taken the underlying cognitive process of the nudge into consideration. However, looking at his examples, this is not clear. According to Saghai (2013, p. 4), a 'TV ad for a pharmaceutical product that lists the side effects read in a monotonous voice while displaying images

of butterflies and happy people' is a not an easily resistible influence because it 'triggers non-deliberative shallow cognitive processes'. He states that 'by providing the mind with enjoyable and attention-grabbing stimuli irrelevant to the audio disclosure its strategy is to confuse the audience through the association of those stimuli with positive effects, rather than the negative effects typically associated with worrisome side effects'. In this case, Saghai deems this intervention to be a behavioural prod due to the cognitive process by which the nudge works, and because using this mechanism confuses the audience.

Saghai changes the criterion in the second example, the "Less Than You Think" campaign. Saghai argues that the "Less Than You Think" campaign qualifies as a nudge. The "Less Than You Think" campaign provides information about binge drinking among students at the University campuses in Montana. The campaign works because it breaks a case of pluralistic ignorance. In essence, the majority of students think that the percentage of students who binge drink is higher than the reality; the campaign provides real information to update beliefs and reduce alcohol consumption. According to Saghai, this intervention is an easily resistible influence, which respects attention-bringing capacities and inhibitory capacities. Strangely, he argues that the same campaign, but using an inflated percentage, would be a behavioural prod. In this case, the misleading information would 'undermine the students' ability to make their own choices' because 'their capacity to inhibit endorsing the belief they are exposed to is weakened when they expect the authority communicating with them to be truthful' (Saghai, 2013, p. 4). Unlike the previous example, in this case, the cognitive process and cognitive effort that students have to make to resist the nudge is the same, whether the information is real or not. Therefore, the criteria that identify this example as a prod are different from the criteria that Saghai uses with the TV pharmaceutical ad.

In the case of the "Less Than You Think" campaign, Saghai finds the intervention to be problematic, not as a result of the process by which it works, but because it involves deceiving. I understand that one may have reservations about whether an administration should use false information to change behaviour. At the same time, I do not see why a fake campaign is more problematic than a real campaign in terms of attention-bringing and inhibitory capacities. As Bovens (2013) argues, whatever the information, it is easy to become aware of the message and the intention of the nudge. In the "Less Than You Think" campaign, irrespective of whether real or false information is provided, we can expect agents to become aware of the nudges and their intention, as in both cases they respect attention bringing capacities.

The point of revising the above classifications and their criteria is twofold. Firstly, it illustrates that nudges are different, and there are shared intuitions about what makes some nudges problematic. Secondly, it shows that the criteria used to operationalise these intuitions vary significantly. The review of the different classifications stresses that not all nudges are intrinsically unproblematic in terms of reflection. However, the classifications do not match, and the same interventions may be good or bad

nudges depending on the classification and criteria that we take as a reference. The fact that the classifications show contradictions implies that while there are interesting contributions about how to classify nudges, there are still no conclusive answers on how to judge the moral acceptability of nudges. Authors seem to favour distinctions that rely on cognitive factors when attempting to tackle the ethical implications of nudges. The underlying idea is that nudges can be harmful to autonomy due to the cognitive mechanisms that they exploit. If we can identify the psychological mechanisms by which nudges work, then we can determine whether a nudge is ethically acceptable or unacceptable. The distinction between System 1 and System 2 has been broadly accepted because it supposedly distinguishes nudges by their underlying cognitive mechanisms. Saghai and Baldwin's classifications also search for answers by looking at cognitive mechanisms. In general, many agree on the fact that the best way to address the ethical debate about nudges would be to correctly distinguish the underlying mechanisms that explain nudges' influence on choices (Grüne-Yanoff, 2016; Grüne-Yanoff & Hertwig, 2016; Hertwig, 2017).

As discussed above, the main worries surrounding the reflection objection are that nudges bypass deliberation, work unconsciously, and are non-transparent (work unnoticed and do not work when noticed). Undoubtedly, an exact account of what psychological mechanisms are triggered by nudges would be useful in order to understand which nudges bypass deliberation and work unconsciously, i.e. which nudges undermine reflection and, therefore, undermine autonomy. However, distinctions based on underlying cognitive mechanisms have significant drawbacks and it is uncertain whether they will be able to resolve the ethical question.

Cognition-based distinctions appear to be challenging for several reasons. Firstly, as previously argued, attempts to classify nudges based on decision-making research are unsatisfactory. There is no consensus on distinguishing which interventions fully or partially bypass deliberation, operate unconsciously, or work through System 1 or System 2. Previously, I have discussed problems in these regard with the available classifications. The distinction between System 1 and System 2 seems to be particularly successful because it appears to be an empirically grounded cognitive-based distinction. However, I have argued that it lacks psychological foundations, uses assumptions, and is based on a theoretical approach with gaps and discrepancies. Similarly, Baldwin's and Saghai's classifications incorporate psychological factors, but their classifications are superficial in terms of identifying underlying cognitive mechanisms. Psychological mechanisms behind nudges have not been perfectly identified, and there is considerable disagreement on how to interpret them. Arguing that some nudges are good, and that others are bad requires judging mechanisms according to specific theories of decision-making in order to state which influence is good and bad for choices. However, the debate within decision-making research is profound, even more so when it comes to classifying influences in normative terms.

Secondly, cognitive-based distinctions are also problematic because of the mismatch between nudges and the mechanisms by which they operate. There is no direct link between a nudge and the mechanism by which it operates. For instance, research suggests that individuals stick to defaults due to the effect of cognitive biases such as inertia bias, status quo bias and loss-aversion (Sunstein, 2015a, 2017a). However, defaults may also work because people interpret them as an implicit recommendation (Smith et al., 2013; Sunstein, 2017a). In this case, because the mechanisms differ, should we understand defaults as an ethically problematic intervention when they trigger status quo bias or loss-aversion, but unproblematic if they work due to being perceived as an implicit recommendation? It is true that the ideal would be to trigger one (or n) specific mechanisms in a controlled and foreseeable way, but at the same time it is hard to discard other mechanisms or unforeseen and undesired effects. Such heterogeneity in the responses makes it even more difficult to establish by which psychological mechanism a nudge has affected a choice.

Finally, there is a related practical problem. Valuing nudges according to their mechanisms seems to suggest that all responses to a specific nudge will be equal; however, responses to the same intervention might vary depending on the context of application, subjects' cognitive capacities, and subjects' previous preferences.

The focus and favouritism for cognitive-based distinctions goes in line with the objections of reflectiveness; it carries the idea that whatever the nudge, its goals and its context, changing behaviour appealing to non-reflective mechanisms is wrong. Distinctions based on cognitive mechanisms try to suggest that some nudges operate through unproblematic mechanisms but still maintain that some others operate through problematic mechanisms and are morally wrong. The classifications based on cognitive criteria would be less problematic if the elements to use to classify interventions were more robust. However, the available cognitive distinctions lack psychological foundations and rely on theoretical and normative arguments that often lack empirical support. Likewise, clear information about the mechanisms that underlie the influence of nudges would not resolve the issue. Mechanisms in themselves are not good or bad; it is the theories used in decision-making research and approaches to decision-making and autonomy that make a mechanism problematic or unproblematic.

The focus on identifying the nudges underlying the psychological mechanisms to resolve ethical issues implicitly incorporates a normative notion of how decision-making should be in order to respect agents' autonomy. Conscious deliberation and awareness of the factors that affect behaviour are crucial to ensuring agents' reflectiveness and that nudges respect agents' autonomy. The following section argues that this approach is limited to addressing the ethical implications of nudges.

### *3.2.2. Reflectiveness and autonomy*

I started the section discussing the objection that some nudges fail on ethical grounds because of the means they use to change behaviour. Many of these objections belong within the reflection objection,

which establishes that nudges are harmful because they impede reflection by bypassing or obstructing deliberation and by working in a manner that is not transparent and cannot be identified by subjects. Previous literature indicates persistent concerns about nudges having the potential to shape behaviour without agents' reflectiveness; this implies that agents are unable to recognise the nudge and the effect that the nudge has on their behaviour. The term "automatic" usually captures these concerns and denotes the potential of nudges to steer individuals from A to B by minimising agents' effort and attention, thus minimising agents' reflectiveness.

The reflection objection uses a specific notion of autonomy, which, I suggest, is restricted to considerations about how nudges affect the decision-making process. Following this idea, a dimension that appears relevant to discussing the implications of nudges on autonomy is the "outcome vs process" dimension. The "outcome dimension" refers to the decision-makers' ability to do what they want depending on their preferences and desires; the "process dimension" relates to the decision-makers' capacity to evaluate options and deliberate.

Chapman (2016) and Engelen (2019) use similar distinctions to discuss the implications of nudges on autonomy and rationality. Chapman (2016) defines outcome autonomy as the 'nonparticipatory linking between motivations and choices', and process autonomy as a participatory process of 'actual participation as a decision-maker'. He uses this distinction to assess the ethical implications of nudges. Similarly, Engelen (2019) establishes a distinction between outcome-oriented and process-oriented rationality. By 'outcome-rationality', he is referring to the outcome of a choice; by 'process-rationality' he is referring to the psychosocial processes that bring about this choice (Engelen, 2019, p. 208). Engelen also uses the distinction to assess the impact of nudges on rationality and value their ethical acceptability. Although Chapman (2016) focuses on autonomy and Engelen (2019) on rationality, in both cases, the outcome dimension refers to the result of choice, and the process dimension refers to the route by which a particular choice comes about. In both cases, these two dimensions are crucial to understanding how nudges affect people's choices and in which ways they can promote or compromise their decision-making and autonomy.

The outcome vs process distinctions help to emphasise the main worries about nudges. In the rationality objection, nudges are problematic because, even though they promote outcome rationality (i.e., steer people towards rational options), they compromise process rationality (i.e., change behaviour by appealing to irrational factors). The reflection objection is also concerned with the issue of process autonomy. Process autonomy is about choices made in the right way. The worry is that some nudges operate in ways that negatively affect the decision-making process and compromise process autonomy by limiting people's reflectiveness.

A priori, the notion of autonomy as reflectiveness appears more plausible than the notion of autonomy as rationality. In this case, individuals do not need to behave rationally and act for rational reasons, the

notion of autonomy as reflectiveness is more relaxed and emphasises that agents have to be conscious of the factors that affect their choices and engage in some reflection to produce a choice. While less demanding, this notion also has some problems. The objections to nudges on the bases that they compromise reflectiveness maintain high standards about how the process of decision-making ought to be. Likewise, they overlook issues arising from the context of choice that might affect the decision-making process and undermine people's outcome autonomy. Different responses to the worries posed by the reflection objection indicate that focusing only on how nudges affect the decision-making process is limited when it comes to assessing their ethical implications on autonomy.

The notion of autonomy as reflectiveness gives a central role to consciousness and awareness in decision-making. Critics of nudges emphasise that agents should engage in conscious deliberation and argue that some nudges appear to be harmful to autonomy because they influence behaviour unconsciously. Likewise, detractors of nudges argue that people ought to be aware of the external influences that affect their choices and stress that some nudges are harmful because they work covertly. These objections, however, overlook the role that unconscious factors play in everyday decisions.

Research suggests that the vast majority of our choices are affected by unconscious influences (Felsen & Reiner, 2015; Newell & Shanks, 2014; Uhlmann, Pizarro, & Bloom, 2008). In general, we act upon many factors that we do not perceive, and we are rarely fully aware of all the factors that affect our decisions (Conly, 2013; Felsen, Castelo, & Reiner, 2013). Therefore, it is unclear why unconscious and covert influences should be labelled as being problematic. At the same time, even if we are unaware of the intervention or relevant psychological mechanisms that bring about a particular choice, we are usually aware of the choices we make and we can reflect on them (Uhlmann et al., 2008). So, even if non-conscious and covert factors affect our decision, in a vast majority of contexts 'we still consider ourselves as the authors of these choices, *post hoc*' (Nagatsu 2015b, p. 489).

A conception of autonomy formulated primarily on the basis that people have to be overtly influenced and aware of the factors that affect their choices is inconsistent with empirical evidence on decision-making (Felsen & Reiner, 2011, 2015). Many factors that affect agents' choices are not directly evident to them and influence their behaviour unconsciously. At the same time, the fact that some nudges might influence behaviour covertly is far from meaning that agents will be completely unaware of their resulting behaviour or choice. They may be unaware of the nudge, or the mechanism, but they are unlikely to be unaware of their own behaviour.

Several authors note that, when assessing the ethical value of nudges, critics make implicit assumptions about how agents' reasons and motivations come about, particularly about how internal and external factors influence their choices. The requirement that agents have to be aware of the factors that affect their decisions is associated with the idea that there are "pure" reasons that guide people's decisions. Implicitly, critics rely on an ideal notion that assumes that there "authentic" internal causes for action

that should be respected and preserved to guide individual decisions (Felsen & Reiner, 2015; Fischer & Lotz, 2014; Schubert, 2015). Detractors of nudges seem to presuppose that internal causes are defined, stable and perfectly distinguishable from external influences; consequently, nudges are external, intentional influences that disturb pure internal reasons and, therefore, undermine people's autonomy. In this schema, for nudges to be normatively acceptable, they must influence behaviour openly. Nudges must be transparent and engage conscious reasoning, so that people are aware of the nudge and can consciously balance their authentic internal reasons with the external influences.

Thaler and Sunstein (2008) argue that choices are always affected by the disposition of options in the context of choice as a fundamental argument in defending the acceptability of nudges. The authors claim that the effect that choice architecture has on behaviour is inevitable, in the sense that individuals' decisions will be affected by other arbitrary factors beyond their control in the absence of nudges. Building on this argument, Thaler and Sunstein (2008) emphasise that opposing nudges is nonsensical, because people are always being nudged, by which they mean always being influenced by non-controlled factors in the choice architecture.

Similarly, Schubert (2015, p.8) argues that, 'when debating the ethics of nudging, we should stop idealising the institutional status quo. Most critics seem to implicitly assume that, before nudges are implemented, people act upon preferences that are somehow 'pure' or 'undistorted' and that nudging then spoils the show by distorting processes of preferences formation'. However, the notion of "pure" preferences does not make conceptual sense (Felsen & Reiner, 2015; Schubert, 2015). People's decisions are highly shaped by their context of choice. In the absence of nudges, subjects' actions are guided by "pure" and "authentic" reasons but by reasons that reflect the influence of external and internal factors and constraints (Schmidt & Engelen, 2020).

The context of choice is full of factors that affect people's choices, and these external influences are rarely neutral. Random influences with no intentionality are a minority. Many companies use nudging techniques to intentionally influence people's decisions, shape their preferences and affect their behaviour. As Schmidt & Engelen (2020, p. 3) argue, 'the choice we often face is not between "intentional" and "unintentional" choice architecture but between different kinds of intentional choice architecture'.

People's choices, values and actions are not the consequence of deliberation based on purely internal motives, but rather the result of a combination of internal and external factors. The notion of autonomy of reflectiveness loses significance in light of the evidence that individuals adjust their preferences and aspirations to their possibilities (Elster, 1983), that the context of choice heavily influences people's behaviour, and the confirmation that many processes, whether conscious, unconscious, intentional or non-intentional, affect individual desires and choices.

Several philosophical notions of autonomy emphasise that individuals have to be able to make the choices they want to make in order to be autonomous, and that these choices should reflect their true preferences and be authentic and consistent with their ‘higher-order desires’ (Bovens, 2009; Dworkin, 1988; Felsen & Reiner, 2011). The arguments under the reflection objection stress that nudges compromise autonomy because they corrupt the formation of preferences and desires, and people’s choices no longer reflect their authentic preferences. However, these objections put too much emphasis on process autonomy and overlook how much people’s choices are shaped by the context and how many factors beyond the agents’ control tend to influence their decisions.

Building on the findings that emphasise the effect that the context of choice has on shaping choices, and the fact that agents struggle to control the factors that affect their behaviour, there is growing recognition of the fact that autonomy requires a combination of internal and external conditions (Mills, 2013, 2015). Proponents of nudges defend that nudges have the potential to work as external, intentional sources of influence that will ensure a better relationship with internal influences, leading to an increase in personal autonomy. Mills (2015) argues that ‘some instances of choice architecture are not only compatible with personal autonomy but can promote it’. The ‘decision-making situation can be designed to protect authentic choices so that it contains options that the individual would choose in a situation free from obstacles (either internal or external)’ (Mills, 2013).

In line with this response, some authors identify cases in which nudges promote outcomes that align with people’s motivations and preferences. Bovens (2009) discusses situations in which nudges are likely to steer people in directions that are consistent with their overall preferences and may improve people’s capacity to make autonomous decisions. He argues that in situations of akrasia, ignorance, inertia, queasiness, and exception, nudges are likely to improve people’s outcome autonomy. In situations of akrasia, cases in which first-order desires compromise second-order desires, nudges can help by structuring choices in ways that promote people’s long-term goals. In instances of ignorance (i.e., situations in which people have limited information or knowledge about a topic), nudges can provide guidance. In the case of inertia (i.e., situations in which people know what they should be doing but fail to do so due to inertia), nudges could break this pattern. In instances of queasiness (i.e., situations in which people do not oppose a choice but avoid making it because it has an emotional cost), nudges could help overcome this cost. In cases of exception (i.e., situations in which people often regret the choices they make), nudges intended to delay a decision could be helpful.

Similarly, Thaler and Sunstein (2008) also discuss situations in which nudges could be particularly beneficial. For instance, nudges that aim to close the intention-action gap (i.e., the gap between what people intend to do and what they actually do) are nudges that promote people’s best interest in line with their second-order desires. Likewise, in cases in which the context tends to obstruct people’s ability to make informed choices, nudges also promote outcome autonomy. For instance, in cases of choice overload (i.e., situations in which individuals face too many options and struggle to process the

information), agents feel overwhelmed and tend to avoid choosing; consequently, a nudge might act as a choice enabler. In contexts in which people lack feedback or in situations in which there is a gap between the causes and consequences of actions, nudges might be helpful in closing this gap. Likewise, in situations that involve a high degree of difficulty or complexity, nudges could be used to promote informed choices or to facilitate choice. Finally, in decisions that people encounter less frequently and have less practice in making, a nudge could be beneficial to providing guidance. The examples given by Bovens (2009) and Thaler and Sunstein provide interesting considerations about the fact that external interference in people's choices can enhance their outcome autonomy.

Conceptions of autonomy that stress that autonomous decisions should be consistent with higher-order desires (e.g., Dworkin 1988), balance a trade-off between letting people reflect on their own choices (process autonomy) and helping them achieve outcomes in line with their higher-order desires (outcome autonomy). There is tension between respecting people's reflectiveness and promoting positive outcomes that aligns with their preferences to promote their autonomy. Excessive attention to how nudges undermine process autonomy might detract from the fact that they can also work to promote outcome autonomy. While paying attention to issues of reflectiveness is necessary, we may miss important variables that could also contribute to promoting autonomy by focussing only on reflectiveness and process autonomy.

The arguments under the reflection objection express valid doubts about how nudges affect the decision-making process and in which ways nudges can compromise people's autonomy. In this section, I have addressed some of these worries. In short, I have argued that nudges cannot generally be rejected on the bases that they undermine reflectiveness. Firstly, nudges differ in how they affect choices, and the assessment of their ethical implications should pay more attention to how different nudges operate. Secondly, the notion of autonomy as reflectiveness is too idealistic. By emphasising that agents need to be aware about nudges, critics implicitly assume that people make decisions based on purely internal reasons. However, given the effect that choice architecture has on people's choices, this does not make conceptual sense. Likewise, the objections on reflection put too much emphasis on how nudges affect process autonomy and tend to overlook how retaining process autonomy might be counterproductive to outcome autonomy and agents' overall welfare. Concerns on reflection are important, but they are not sufficient to discard the use of nudges in policymaking.

Detractors of nudges point out that the inevitability of choice architecture is not sufficient to grant nudges free licence; a non-intentional, random influence does not have the same implications as a deliberate intervention aimed at steering people's behaviour towards a specific end. In line with critics of nudges, I understand that the inevitability of choice architecture is not sufficient for the general acceptability of nudges. Covert influences cannot be applied acritically just because other non-intentional covert influences generally affect behaviour. The intention and rationale of the intervention have to be discussed if an administration is planning on using nudges. However, I do think that

confirming the fact that many non-conscious and non-transparent factors affect choices relaxes objections to reflectiveness, particularly when it comes to the effect of nudges on process autonomy.

#### **4. The moral acceptability of nudges: a combination of three aspects**

So far, this chapter has discussed some of the central arguments against nudges. Sections 2 and 3 respond to the principal objections and argue that nudges are not intrinsically problematic as a result of either their aims or their mechanisms. While the objections express legitimate worries, they are not conclusive and can be addressed.

Although nudges are not intrinsically problematic, they are not unproblematic; their implementation raises ethical questions that should be addressed. In this sense, we should explore what makes a nudge morally acceptable and which factors are pertinent when discussing their ethical implications. The discussion in the previous sections identified several issues that are relevant when it comes to judging the moral acceptability of nudges. The debate is focused on three themes in particular: the type of nudge, the context of use, and the aim it promotes.

Firstly, the aim of the nudge is a crucial factor. The moral assessment of nudges should include considerations about whether a nudge is being used for good purposes. Secondly, the type of nudge is also an essential element. As argued in section 3, several authors argue that nudge interventions operate in different ways, and this is relevant when it comes to judging their normative acceptability. This is less worrisome for those nudges that rely on deliberation, engage reflection and are noticeable, and is more of a concern for those nudges that trigger unconscious processes, work covertly and rely on automatic responses. Thirdly, the context in which the nudge is applied is also a crucial factor. The decision-making context of a nudge involves two different factors: (i) the nature of the decision-making, and (ii) the issue at stake. The nature of the decision-making has a bearing on the degree of reflectiveness that agents have over decisions. The issue at stake is related to the significance of the choices in specific domains. The implementation of nudges seems to be less problematic in contexts in which people lack reflectiveness and self-command, and the context of choice profoundly shapes their choices. By contrast, nudges seem to be more problematic in contexts in which choice architecture has less effect on people's decision. Likewise, the use of nudges is less daunting in contexts in which people are indifferent about their choices. However, it faces more objections in domains with significance for character formation or that have moral implications. The moral acceptability of nudges is a combination of the three factors: the aim of the nudge, the type of nudge, and the context of use. The sections that follow look at how these three aspects affect the moral implications of nudges and how they relate to each another.

#### **4.1. The aim**

Policy instruments must be used for relevant and ethically justified policy goals. When applying nudges, policymakers must explain why and where it is necessary to interfere in people's choices, how nudges can help with specific problems, what the rationale is for justifying their implementation and what the positive and negative implications are that they potentially promote. These questions, however, do not apply to nudges alone. Policymakers should address these doubts for any intervention, clarify the rationale for intervention and address the implications of implementing them. As a result, nudges face the same issues as other behavioural change interventions as far as aims are concerned.

However, worries persist about nudges being easier to use to promote illicit ends because they can be applied with limited democratic control. According to Lepenies & Malecka (2015), nudges are often applied without legislation and may evade political control mechanisms, particularly if they are implemented by an external "nudge unit" that lacks democratic accountability (Lepenies & Malecka, 2019). Pérez (2013) also notes that it is more challenging to publicly check and monitor nudges, whereas more coercive measures such as sanctions and incentives are easier to control. To some, these worries are aggravated by the ways in which nudges influence choices. Lepenies & Malecka (2015) also stress that because nudges evade political control and exploit non-deliberative mechanisms, their use undermines the public discussion and reflection over the goals they promote.

Every aim and tool that the governments use should be publicly discussed and scrutinised. The implementation of nudges should be debated, explained and used to pursue relevant and ethically justified policy goals. Lepenies & Malecka (2015) argue that interventions that involve rational persuasion should be preferred because, in these cases, policymakers have to rely on reason-giving and people are more openly conscious and involved in the discussion of ideas and goals that such interventions promote. However, educational and persuasion campaigns are also applied without legislation. Although these interventions rely on rational persuasion, they can also be challenging to monitor and can degenerate into indoctrination and other forms of manipulation (Pérez, 2013). In the same way that it would be an exaggeration to argue that governments generally use these tools for unethical aims, it is also an overstatement to suggest that the use of nudges favours the promotion of illicit ends.

To be morally acceptable nudges need to promote ethically justified ends. As policy tools, their use must be subjected to regular control mechanisms, taking the outcomes policymakers want to pursue through them into consideration and discussing their implementation through traditional democratic means.

## 4.2. The type of nudge

The type of nudge is a relevant factor when it comes to judging the moral implications of nudges. Problems of conceptualisation underpin the discussion on types of nudges and ethical value. As discussed in section 3.1, it is tricky to categorise nudges as interventions that exploit irrationality due to competing approaches to rationality within decision-making research. Likewise, section 3.2, has also established that, beyond the debate on how nudges relate to rationality, nudge interventions operate in different ways, and considering these differences is crucial when it comes to assessing their ethical acceptability.

Section 3.2.1 presented some of the attempts made to classify nudges and discuss their ethical implications. The majority of classifications assume that knowing the underlying cognitive mechanisms by which nudges operate informs the normative implications of different nudges, particularly in terms of autonomy. The essential idea is that we can distinguish unproblematic nudges (i.e., nudges that engage reflection and are noticeable) from more problematic nudges (i.e., nudges that bypass deliberation, undermine cognitive awareness and benefit from an automatic response).

However, in section 3.2.1, I addressed the difficulties of judging the ethical validity of nudges by looking at their underlying psychological mechanisms. Cognitive-based distinctions struggle to provide conclusive answers for several reasons. Firstly, there are problems in identifying the mechanism by which nudges work due to disagreements and limitations in decision-making research. Likewise, given that what we actually know about decision-making is limited, assessing cognitive mechanisms often implies making assumptions about the mental process at work and incorporating normative notions of behaviour and rationality. Secondly, there are practical limitations, given that the same nudges may work through different mechanisms, and their implementation might produce heterogenic responses. Both issues make it challenging to distinguish nudges using cognitive criteria. Finally, cognitive-based distinctions implicitly assume that nudges that engage conscious awareness are morally superior and should be preferred. However, while this feature is relevant, it should not be the decisive criterion when it comes to judging the moral permissibility of nudges.

At the core of the debate on different types of nudges is the likelihood that different interventions affect choices without awareness. The debate on awareness involves three different considerations: i) whether agents are aware of the nudge intervention, ii) whether agents are aware of the mechanism that it triggers, and iii) whether agents are aware of the outcome that it produces. Critics sometimes conflict these dimensions.

To assess whether nudges have the potential to steer behaviour without awareness; rather than focusing on the psychological mechanisms that nudges engage, we should focus on features of design and implementation as behavioural change interventions. Given the limitations of cognitive-based distinctions, I argue that the potential of nudges to change behaviour with limited awareness (of the

nudge, the mechanism, and the outcome) is related to features of nudges and elements about their design and implementation. Three specific features appear relevant when looking at the capacity of nudges to produce an automatic response. Firstly, whether nudges expect an active or passive response; secondly, whether the intervention is transparent or non-transparent; and thirdly, whether the intervention is designed to affect behaviour within a specific choice situation or to affect behaviour across different choice situations. These three features outline how a nudge will operate and affect behaviour and the likelihood that such features will reduce awareness of the nudge, the mechanisms and the outcome.

- *Active vs Passive response*

In the normative discussion on nudges, several authors suggest that interventions differ in how they engage deliberation and use this characteristic as a way of assessing the ethical acceptability of nudges. Previously, I have argued against the idea of nudges as interventions that exploit failures of rationality and always work by bypassing or obstructing deliberation. Likewise, I suggested that looking into the cognitive mechanisms that underlie a nudge's effect might not be as promising as it appears in terms of discerning interventions and addressing their normative implications. Distinguishing how different nudges engage deliberation varies, depending on the theory of decision-making adopted (either rationality, reflection or System 2) and, therefore, it is challenging to agree on how different nudge interventions relate to deliberation.

In an alternative approach, Hansen & Jespersen (2013), present an epistemic distinction between Type 1 and Type 2 nudges that is not based on the cognitive mechanisms at work. In their classification, Type 1 nudges change behaviour without involving reflective thinking; Type 2 nudges involve reflective thinking, and mainly influence behaviour resulting from some degree of deliberation. They argue that Type 1 nudges directly affect choices, whereas Type 2 nudges affect the premises of choice. Both types of nudges can influence behaviour via automatic models of thinking, such as automatic process, unconscious factors, or System 1. However, Type 2 nudges influence reflective thinking and Type 1 nudges influence automatic thinking.

Their classification encapsulates a fundamental question: whether interventions require an active or passive reaction from the nudgees to succeed. Nudges that seek a passive response shape choice architecture in ways that allow the nudge to change behaviour with or without agents' effort and deliberation, i.e., without an agent's active contribution. Nudges that seek an active response require reflection from nudgees to work and do not work if there is no reaction about the nudge.

To illustrate this point, interventions such as "the cafeteria nudge", reducing the size of food portions to reduce calorie intake (Wansink et al., 2005; Wansink & Cashman, 2006) and placing healthy foods in more accessible areas (Dayan & Bar-Hillel, 2011; Keller, 2014) are interventions that work with a passive response. These interventions reduce the deliberation effort to a minimum, affect choices directly and work if agents have a passive response. Similarly, default rules also require minimal

deliberation. Although they may work depending on different factors, for instance, implicit endorsement, status quo bias, loss aversion, or lack of effort (Smith et al., 2013; Sunstein, 2017a), if the agent does not actively react to the default, the intervention is effective.

By contrast, nudges such as warnings and reminders, the communication of social norms, and framing effects are interventions that affect the premises of a decision and are designed to activate some degree of reflection. These interventions require an active response, act as added inputs to lead choices towards a particular outcome, and their effect on behaviour is indirect. In these cases, if there is no active reaction to the nudge, the nudge will be ineffective. It could still be argued that these interventions exploit System 1 and inhibit reflection. Likewise, it could be argued that some of these interventions, such as framing or the communication of social norms, exploit a non-rational, response. However, regardless of the mechanisms in play, these interventions do not work without an active response from the nudgee.

Establishing whether nudges require an active or passive response does not require judging the content or mechanisms of the response. The crucial factor is that nudges that expect a passive response may work even when subjects do not react to the intervention, whereas nudges that expect an active response only work if agents respond to the nudge.

- *Transparency*

A second crucial element in the discussion on the means of nudges is their transparency. Nudges have been accused of working without transparency and being difficult to perceive as behavioural change interventions. As previously argued, the fact that nudges are often not directly evident to the nudgees is taken as a feature with negative ethical implications.

However, in section 3.2.1, I presented two accounts that distinguish between transparent and non-transparent nudges using design criteria. In both accounts, transparent nudges are defined as noticeable interventions, and non-transparent nudges are interventions that are difficult to notice and that make it hard for subjects to recognise the nudge and the intention behind its use. Following these two accounts, I also argue that it is possible to distinguish between transparent and non-transparent nudges. Transparent nudges are interventions that are easy to perceive. Agents can quickly identify the nudge and understand that it is an intervention to shape their behaviour in a particular direction. By contrast, non-transparent nudges are interventions that are difficult to perceive, and for which it is challenging to become aware of the intention behind their use.

To illustrate this point, interventions that use messages, feedback, information and that communicate social norms, even those that seek an emotional response, are noticeable. Likewise, warnings (texts or images), reminders and interventions to prompt choices are also noticeable nudges. These interventions only work if the subjects perceive them. Likewise, in all these cases, agents can easily recognise the intervention and understand the direction in which the nudges want them to go. In this sense, even if

the effect of the nudge is strong, for instance eliciting a strong negative emotion towards an option, it is easy for agents to become aware of the intervention and its intention (Bovens, 2013).

By contrast, interventions such as reorganising environments, default rules and framing devices are more difficult to perceive as behavioural change interventions. In these cases, nudges may work without agents' perception of the intervention, and without agents' perception of the fact that their behaviour has been influenced in a particular direction. When covert influences are at play, subjects are not aware of the factor that is shaping their choices, and this triggers doubts about their ethical implications.

Bovens (2009) argues that some nudges 'work better in the dark' and maintains that nudges cannot be disclosed to the subjects because, once people find out about the nudge and about how it is supposed to influence their choices, they tend to be less inclined to follow it. Psychological research on priming effects indicates that when people are made aware of how their own choices were affected by an unconscious clue, they tend to change or correct their decision-making (Uhlmann et al., 2008). Likewise, research on psychological reactance also stresses that people assign intrinsic value to making choices and, therefore, reject external influences. Some studies indicate that people prefer to choose even at the expense of the outcome (Bobadilla-Suarez, Sunstein, & Sharot, 2017); people like to retain control over their choices and dislike influences that undermine (or that are perceived as undermining) their freedom of choice. Consistent with Bovens' argument, these findings indicate that disclosing non-transparent nudges may diminish their effectiveness, people may perceive the nudge as a threat to their freedom of choice and dismiss the nudge or choose the opposite direction.

However, empirical research indicates that disclosing nudges does not necessarily decrease their effectiveness. In a field experiment in a train station snack shop, Kroese et al. (2016) found that notifying people of the purpose of the nudge did not have any impact on the effectiveness of the intervention. The intervention found that placing healthy snacks near the cash register strongly impacted on customer choices. Moreover, adding a sign with the message, 'We help you make a healthy choice' did not have any effect on sales. Similarly, in a laboratory experiment, Loewenstein et al. (2015) found that disclosing the effect of defaults did not weaken their effectiveness to promote a pro-social goal. Subjects that were informed of the default and were later allowed to change their choice did not modify it. Finally, in a laboratory experiment, Bruns et al. (2018) investigated how informing decision-makers about (a) the potential influence and (b) the purpose of the nudge affected the effectiveness of nudges. The authors found no effect of transparency on the effectiveness of pro-environmental defaults. All these findings suggest that in situations where there are ethical concerns about nudges working covertly, their use can be disclosed, and nudges may still be effective.

In conclusion, the noticeability of nudges varies across interventions. Some nudges are easy to identify as behavioural change interventions and agents understand that their behaviour is being influenced in a specific direction. Some other interventions are more hidden, unnoticeable and both the intervention

and its intention are not easy to perceive. Non-transparent nudges are more likely to produce an automatic response and work without agents being aware of the nudge. Non-transparent nudges work covertly, which implies that agents are not aware of the nudge. However, unnoticeability can be corrected. In some cases, disclosing information about a nudge (the intervention, its influence or its purpose) may resolve ethical worries, without undermining the nudge's effectiveness. In this sense, non-transparency should not be taken as an intrinsic nudge attribute but rather as a malleable feature in the design and implementation of the nudge.

- *Choice situation*

Nudges are designed to alter a choice or behaviour in a particular environment. Their implementation is usually attached to a specific choice set. Most nudges work because they are attached to the choice situation in which a particular choice or behaviour takes place. However, nudges vary in terms of the choice situations in which they are effective. Some nudges have to be attached to the choice execution context. By contrast, other nudges do not have to be attached to the immediate choice set to work and may influence choices across different choice situations<sup>7</sup>.

To illustrate this point, let us take as an example of two nudges that target driving behaviour with the intention of reducing speeding: the use of increasingly narrow lanes on roads and the use of emotional messages on roadside panels. The use of increasingly narrow lanes on roads to reduce speeding is a nudge that only works in a particular choice situation. The nudge is only valid when agents are driving on those particular roads. By contrast, the communication of emotional messages on roadside panels is a nudge that does not need to be attached to the actual choice set. The nudge is intended to affect the same behaviour (speeding), but the messages have the potential to affect behaviour beyond the immediate choice situation in which the nudge is applied.

Nudge interventions that have to be attached to a choice situation include default options, the re-organisation of a physical environment, the alteration of the size of products and some warnings. These interventions are designed to affect choices in specific situations and only work when attached to the choice execution context. By contrast, nudge interventions that are not attached to a choice execution include the communication of social norms, or the communication of moral and emotional messages, the use of reminders, some types of warnings, and the provision of feedback. These interventions are not attached to the choice execution and potentially may affect behaviour across different contexts.

Nudges that are attached to a choice execution are more likely to affect behaviour, with agents potentially not being aware of the nudge, the mechanism and the outcome; they are more likely to produce an automatic response. When the nudge is attached to the immediate choice execution, it can

---

<sup>7</sup>To the best of my knowledge, this characteristic is not considered in previous literature. In a footnote, Bovens (2009) briefly mentions that this trait could be used as a feature to distinguish nudges “attached to a choice situation” from what could be a similar intervention, but not nudges “not attached to a choice situation”.

affect behaviour without agents being aware of the nudge. When the nudge is not attached to the context of choice execution, it is less likely to affect behaviour without agents being aware of the nudge.

The combination of the three features: type of response, transparency and choice situation, outlines the likelihood of nudges affecting agents' behaviour by reducing conscious and reflective response. Non-transparent nudges are covert influences. Nudges that seek a passive response want to minimise agents' effort and attention. Both features make nudges able to produce an automatic response, in which agents are unaware of the nudge, its influences and, in some cases, its outcome. The automatic response can only be effective if nudges are attached to a choice situation.

The choice situation in which nudges work is also a relevant feature when it comes to understanding how nudges can affect long-term behavioural change.

- *Classification of nudges.*

Let us now look at how some of the most common examples of nudges relate to the three features.

Default options are nudges that seek a passive response, where, if the agent does not actively communicate that they wish to opt out of the default, the pre-selected course of action takes place. The crucial point here is that, regardless of the mechanisms that are at work, defaults are effective with no active reaction over the issue at stake. In terms of transparency, defaults are a non-transparent nudge. When the default is not disclosed, it is difficult for decision-makers to perceive the intervention or its aim. Likewise, even in cases in which they perceive that a form has a pre-selected option, it is still difficult for individuals to see this as an intentional intervention to shape their behaviour. Moreover, defaults may work even in cases in which they are unperceived. Finally, default options need to be attached to the specific choice execution and are only effective at the point of the choice.

The combination of these features makes defaults a form of nudge that is likely to affect agents' behaviour, with them having limited awareness of the nudge and no awareness of the mechanism. Likewise, in some contexts, defaults may change behaviour without agents being aware of the outcome. For instance, if a way of deciding between A and B also includes a choice over a secondary subject, between C and D, with C being pre-selected, agents could end up selecting C without noticing. Under conditions of cognitive load, if agents are fixated with the decision between A and B, they may overlook the choice between C and D and ignore or be unaware of the outcome of their choice. Cases that follow this structure are common on online purchasing forms. For instance, there are often pre-selected extra options on travel insurance, extra baggage options or seat selection on online forms for purchasing plane tickets. The selection of these options is not fully disclosed and tends to be challenging to appreciate. In these cases, people may ignore or be unaware of the fact that they have purchased other elements in the process of buying the tickets. Similarly, registration forms sometimes incorporate pre-selected options to sign up for secondary courses, donations to charity or gym memberships. In these contexts, defaults can also mean that people are unaware of the outcome of choice.

**Table 5. Features of nudges interventions.**

<b>Nudge</b>	<b>Type of response</b>	<b>Transparency</b>	<b>Choice situation</b>
Default options	Passive	Non-transparent	Attached to the choice execution
Organisation of the physical environment	Passive	Non-transparent	Attached to the choice execution
Alteration of the size of the portion	Passive	Non-transparent	Attached to the choice execution
Framing effects	Active	Non-transparent	Usually attached to the choice execution
Labels	Active	Transparent	Attached to the choice execution
Reminders	Active	Transparent	Attached to the choice execution
Warnings	Active	Transparent	Attached to the choice execution
Pre-commitment strategies	Active	Transparent	Attached to the choice execution
Use of social norms	Active	Transparent	Not attached to the choice execution

The re-organisation of the psychical environment shares features with defaults. Firstly, it is a form of nudge that expects a passive response; the nudge is effective with no effort or active reaction from subjects. Secondly, it is a non-transparent nudge. If the intervention is not disclosed, people do not expect the intervention, and it is difficult to notice. Thirdly, it is a nudge that only works at the point of choice and has to be attached to the choice execution context. The re-organisation of the psychical environment is likely to produce choices without awareness of the nudge and the mechanism because of these characteristics. People do not perceive the intervention and do not know how the context of choice influenced their behaviour. By contrast, these types of interventions are unlikely to work without awareness of the outcome, since it is reasonable to think that people will be able to notice what they are eating.

The alteration of the size of the portion of food and drinks has similar characteristics. It is a nudge that expects a passive response and is effective if agents do not actively decide to get more product. It is also a non-transparent intervention, meaning that, if the nudge is not disclosed, it will be difficult for subjects to perceive the intervention. Finally, the nudge only works in the immediate choice execution context. The nudge may affect behaviour without agents perceiving the nudge and the mechanism that it is triggering because of these characteristics. Likewise, some research indicates that this nudge may even change the outcome without agents perceiving that they are eating less (Wansink & Cashman, 2006).

Framing effects require an active response. Framing effects alter how information is presented and are aimed at triggering decision-making biases. However, people have to react to the information and make a choice in response; if there is no active choice over the issue at stake, the frame will not work. Research on framing effects indicates that when people are faced with framed information in a foreign language, the frame is less effective because people engage more attention, which may neutralise the effect of the frame. These findings reinforce the fact that framing effects require an active response. This characteristic makes framing effects unlikely when it comes to reducing agents' awareness of the outcome of their choice.

In terms of transparency, framing effects are a bit difficult to classify. In general, framing effects are non-transparent. When perceiving framed information, it is unlikely that agents will recognise the frame. However, the non-transparency of framing effects depends on the context. For instance, public and private campaigns that describe information in specific ways to make a choice more attractive are using framing effects. However, the campaigns are transparent and intentional. In these cases, while agents might not perceive precisely how the frame is affecting their choices, they can easily understand the purpose of the campaigns. People might be unaware of the mechanism that is affecting their choices but perceive the intervention (the public campaigns). By contrast, when an authoritative person, such as a doctor, explains probabilities to a patient using framing effects, the framing effect is challenging to identify, and agents are also unlikely to perceive the recommendations of these people as an intentional intervention. In this context, the intervention is non-transparent. There is no evidence on how people react when framing effects are disclosed. It would be interesting to see how people perceive their effect on decision-making and whether disclosing the frame reduces its effectiveness.

Finally, the distinction between public campaigns that use framing effects and the use of framing effects in single decision contexts is also relevant when it comes to discussing how framing effects are applied in different choice situations. In the first case, a public or private campaign that uses framing can influence choices in a different environment and does not need to be attached to the choice execution context to work. In the second case, the framing only works if it is done closer to the time and place of choice. Because framing effects are not transparent, they are likely to affect behaviour without agents being aware of the nudge and the mechanism that triggers the response. By contrast, because they require some type of active reaction, they are unlikely to reduce agents' awareness of the outcome of the choice. Even if people ignore or are unaware of the intervention and the mechanism that influenced their choices, they will be fully aware of the choice they make.

The use of labels is a nudge that expects an active response. Labels have been used to display information about the number of calories in products and options on menus, and to provide information about and promote eco-friendly purchases of motor vehicles, appliances and light bulbs (Borgmeier & Westenhoefer, 2009; Newell et al., 2013). The information on labels can vary a lot and use different formats to make the label simpler and easier to understand. For instance, some calorie labels use full

information on calorie intake, while others simplify the intervention by employing the traffic light colour system or describe the amount of physical activity required to burn off the calories (Bleich et al., 2012). Some labels require more attention and mental processing, while others are simpler and can help reduce the effort to understand them. Regardless of the format of the information, agents have to process the label and actively react to it; the nudge requires an active response. In terms of transparency, labels are transparent, but the nudge will only work if people perceive the label and understand its intention. Finally, to be useful labels, have to be implemented at the point of choice, either on the product or in the surroundings, and will work only when attached to a specific choice execution situation. For this reason, labels are unlikely to affect behaviour without people being aware of the nudge and the outcome.

The use of reminders is a nudge that requires an active response to be effective. Its implementation is transparent; subjects notice the nudge and can easily understand its intention. Finally, in terms of choice execution, reminders work best when they are close to the point of choice. Similarly, prompted choices in which people are asked how they feel about making a choice, pre-commitment strategies based on which people commit to a specific course of action, and interventions that select implementation intentions by asking questions about future behaviour are interventions that work with the same characteristics. All of these nudges require an active response, are transparent and need to be applied closer to the time of the choice. These nudges are unlikely to change behaviour without agents being aware of the nudge, the mechanism and the outcome as a result of these characteristics.

Finally, the use of social norms is a nudge that requires an active response, is transparent and can be applied both in the context of choice execution and to affect behaviour in different contexts. Regardless of the mechanism by which the communication of social norms affects behaviour, agents need to actively modify their behaviour for it to be effective. The nudge is transparent because it is easy to identify as a behavioural change intervention, and the intention of the intervention is easy to grasp. Finally, in terms of choice execution, the use of social norms can be both present at the point of choice, but also not attached to the choice execution. In some cases, the message is present in the immediate context of choice and works to modify specific behaviour. In other examples, the use of social norms can be useful when it comes to modifying several choices. For instance, The Home Energy Report by OPOWER (Allcott, 2011), which compares household consumption in a neighbourhood using social norms, is independent of a particular choice execution. The information may affect more than one choice execution: for instance, switching off the lights, reducing the water consumption, opening the windows instead of using the air conditioning, etc.

#### **4.3. The context of choice**

Ethical considerations about nudges cannot be made without paying attention to the context of choice in which the nudge is implemented. I use the expression “context of choice” to represent two factors: (i) the nature of the decision-making, i.e., how decisions are made in particular contexts, and (ii) the

issue at stake, i.e. the domain and type of behaviour that the nudges are targeting. Both elements are relevant when it comes to judging the moral permissibility of nudges. The implications of nudges differ depending on how decisions are made, and what they are about.

In the previous section, I suggested that nudge interventions should be distinguished according to three factors: active vs passive response, transparency, and choice situation. I argue that nudges that require a passive response, non-transparent nudges and nudges that have both features may change behaviour without agents being aware of the nudge or being aware of the influence that affected their choices. Likewise, in some contexts, nudges with both features may change behaviour without agents being aware of the outcome. Nudges with one or both features reduce reflectiveness, may compromise people's process autonomy and push individuals towards options that are inconsistent with their desires and preferences.

In line with the above-mentioned debate, many authors would argue that these types of nudges are morally problematic and should be rejected. Transparent nudges and nudges that require an active response should be prioritised because they involve conscious reflection and people know which factors affected their choices. However, responding to objections related to the reflection objection, I argue that the notion of autonomy as reflectiveness places too much emphasis on awareness and conscious deliberation, disregarding the fact that decisions are often shaped by non-conscious factors, and by elements that we do not control. Likewise, the notion overlooks the fact that, in specific contexts, prioritising process autonomy undermines outcome autonomy and welfare.

Making conscious and reflective decisions is essential; it is a relevant part of the pursuit of autonomous choices and important when it comes to developing agents' agency and character. However, individuals' ability and willingness to engage in reflective thinking are inconsistent.

Firstly, individuals' ability to engage in conscious and reflective decision-making depends on the context of choice. In well-known contexts, it is easy to engage reflection and make decisions about choices that we encounter frequently that involve simple options and that allow time. However, when individuals face unknown factors and many options, or do so in conditions of stress, time-pressure or cognitive load, engaging reflection is challenging and may be counterproductive to decision-making. Therefore, it is essential to value the nature of the choice that the nudge is triggering in order to value the ethical acceptability of nudges.

Secondly, agents' willingness to engage in reflective decision-making is not equally relevant across domains. Not all choices have the same significance and value. Some decisions require agents' full attention and reflectiveness because the issue at stake is relevant; in these cases, nudges that reduce awareness and seek an automatic response might raise ethical doubts. By contrast, some other issues are trivial, less critical, or there is a consensus over what is the best course of action to take. In such cases, giving up freedom of choice may be less of a problem.

- *The nature of decision-making*

The argument that nudges compromise people's decision-making implicitly assumes that agents enjoy freedom of choice before the nudge and that no other factors affect their process autonomy. However, as previously argued, research indicates that choices are heavily shaped by the context of choice and by the influence of intentional and unintentional factors that are often not directly evident to the agents. In specific contexts, contextual conditions shape choices and tend to compromise people's decisions. In such cases, nudging might have more positive than negative implications.

As previously mentioned, in '*Nudge*', Thaler and Sunstein (2008) argue that in contexts of choice overload, lack of feedback, complex decisions and decision that have separate causes and consequences, nudges are advantageous when it comes to promoting positive outcomes. The use of non-transparent and passive nudges in such contexts is less of a problem because agents' decisions were already negatively shaped by the context of choice and were beyond the agents' control.

Similarly, Fischer & Lotz (2014) develop a taxonomy of nudge interventions that takes into account the nature of the decisions being nudged. Their taxonomy establishes that nudges that affect unintended behaviour are unproblematic, even if their effect on behaviour is covert and triggers behavioural change on an unconscious level. The crucial aspect is that nudging unintentional choices is not harmful because agents' choices are a result of contextual factors and do not reflect their preferences over the issue at stake.

To resolve issues about the ethical implications of nudges, we need to do more research into the contexts in which people's decisions tend to be more sensitive to behavioural biases, more vulnerable to the influence of contextual factors and less guided by reflectiveness and conscious deliberation. A number of contributions have already identified decisions and situations that undermine agents' capacity to make good, reflective and conscious choices by themselves.

Research on poverty and decision-making has established that a scarcity of material undermines people's cognitive function (Mani et al., 2013). People living in poverty are continually making complex decisions in highly restrictive contexts. They tend to develop tunnel vision through which they focus on meeting their immediate material needs and tend to neglect long-term goals. Conditions of limited material resources make it impossible to defer the acquisition of immediate needs, which is often to the detriment of more beneficial long-term consequences.

Authors suggest that policymakers should be aware of these findings and re-think the design and implementation of programmes and policies to reduce poverty in both developed and developing countries (Curchin, 2017; Datta & Mullainathan, 2014). In many welfare states and developed countries, claiming social benefits, particularly those aimed at the poor, is a time-consuming process, about which people have limited information. Moreover, it often requires complex forms to be filled in and the entire process has to be updated periodically. The process often involves cognitive costs, which

leads to non-take-up of welfare benefit programmes and undermines their effectiveness, thereby unwittingly exacerbating inequalities. In these situations, nudges that simplify the forms and procedures to access social benefits, or that automatically enrol those who meet the requirements will be highly beneficial and not unethical.

Similarly, research indicates that individuals are highly sensitive to decision biases in choices involving financial decisions and tax decisions. Financial issues are a complex area, education strategies are lacking or ineffective, and agents have no incentives for gathering complete information in order to understand the operation of financial products and the tax system. In these complex domains, nudges might be useful even if they undermine reflectiveness, particularly those that are more automatic (Thaler & Sunstein, 2008).

There is growing recognition that the context of choice shapes individual decisions in many domains and that changing behaviour for good requires re-designing options and environments. Previous literature on obesogenic environments, for example, studies the effect of the psychical and social environment, unhealthy food choices and sedentary behaviour. The focus is interesting because responsibility for being overweight and obesity is not laid at the door of individuals, but on the context in which they make choices (Lipek et al., 2015). Looking at the effect of the environment is also relevant when it comes to studying and reducing other health risk behaviours such as alcohol and tobacco consumption. Studying purchasing behaviour and behaviour in areas that thrive on reducing self-control such as gambling and addiction is also relevant.

To look at the ethical implications of implementing nudges it is crucial to examining the areas and domains in which or the material conditions under which individuals are more capable of engaging reflection, and the areas in which they are more dependent on clues and factors from the environment or where they are under cognitive fatigue or stress. Addressing these aspects is also relevant when it comes to protecting individuals better from private nudging.

In general, it cannot be assumed that all choices are made equally. Firstly, when addressing the ethical implication of nudges, we have to pair the nudge (and its specific features) with the decision that it is targeting. When non-transparent nudges and nudges that seek a passive response address unintended or automatic decisions and decisions made in contexts of cognitive pressure and complexity, overriding reflectiveness might not undermine autonomy but promote it. Secondly, we need more empirical evidence about how and in which domains decision-making capacities are weakened due to external factors and discuss how nudges can contribute to correct these situations.

- *The issue at stake*

The worries about nudges being harmful to decision-making and autonomy stem from worries about policymakers interfering in people's choices and undermining their capacity to choose as they want. The implicit idea is that legitimate behavioural change should always be achieved by people's free

choices and should be the result of reflective and conscious decision-making. In this sense, while non-transparent nudges and nudges that seek a passive response might be useful when it comes to helping individuals make good choices, and might be useful for promoting positive societal goals, their use may be harmful because they discourage agents from engaging in active choice (Schubert, 2015).

Choices have intrinsic value; it is through choices that we can define an individual project, construct character and identity and develop a notion of what to value (Conly, 2016; Schubert, 2015). Active and reflective choices (i.e., the ones that agents can identify and rationalise), are valuable for building character. Conly (2016) argues that choices have intrinsic subjective and objective values. On the one hand, making choices allows individuals to shape their personal lives, and to exercise their agency and sense of individuality. On the other, active choices allow individuals to evaluate options, decide according to their own judgment and build their capacity to make good choices for themselves. In this sense, the worry is that nudges that seek an automatic response might undermine the agency and ability of individuals to choose and learn from these choices. However, two additional factors should be taken into consideration.

Firstly, choices are important, but not all choices are equally relevant. In the debate on the ethics of nudges, critics tend to overlook the issue at stake. Conly (2016) argues that ‘the ability to choose means a great deal to us in some contexts, and in others, it means very little’ (Conly, 2016). In this sense, considering the behaviour that is being nudged is essential for assessing the ethical implications of nudges.

It is possible to identify life domains in which agency matters and others in which the effect of a decision on our personality, individuality and capacity to make choices is far less important. Accordingly, there are several cases in which the implementation of nudges can be assumed to be unproblematic. Fischer & Lotz (2014), for instance, argue that nudges that exploit agents’ indifference to alternatives are unproblematic, because they change behaviour in areas in which people do not feel the need to express individual preferences. In these cases, it can be assumed that people will accept the nudges and their outcomes. Similarly, Engelen (2019) argues that many nudges affect trivial choices or promote outcomes that can be safely assumed to be in line with people preferences. In such cases, the use of nudges is inoffensive. For instance, he argues that well-designed stoves, flies in urinals, handles on doors and lines on roads, promote outcomes for which it is fairly clear and uncontroversial what people want. In such cases, there are no heated debates or worrying doubts about what people want. Such nudges are not imposing another’s will on nudgees or substituting their values’ (Engelen, 2019, p. 216).

In general, there is agreement on the fact that in domains in which choices are unimportant, or in which it can be inferred that people will approve nudges, their implementation is harmless. Of course, in some cases, what people perceived as relevant or irrelevant is imprecise and difficult to identify and there may be cases in which nudges do compromise people’s goals and preferences. However, the point is

that the issue at stake is a relevant factor when it comes to addressing concerns about nudges. There are some ideas on domains in which the implementation of nudges is uncontroversial.

By contrast, there are choices and domains in which nudge implementation raises many ethical objections. For instance, one of the most controversial nudges is the use of defaults to increase organ donation. In the discussion, the problem with using defaults to promote organ donation tends to concentrate on the fact that defaults undermine reflectiveness, work covertly and may change behaviour without agents being aware of the outcome. However, these worries are not related to the type of nudge, but with it being used to promote organ donation. Sunstein argues that presumed consent in organ donation could be seen as an intrusion on autonomy and a *pro tanto* wrong. Yet, the reason behind these worries is not about how defaults work, but about the fact that, ‘unless they have given their explicit consent, people should not lose parts of their own body (even at the time of death)’ (Sunstein, 2016, p. 1). Organ donation is an issue where we can expect people to demonstrate deep moral views and which has implications about life and individual agency and character. Therefore, a default would not be the best option to induce behavioural change.

In conclusion, there are decisions for which people have no pre-defined<sup>8</sup> preferences and involve important actions and decisions. Nudges that seek an automatic response might not be the best option when the issue at stake has moral implications and is perceived as a significant domain for building character or confronts social values or social norms. In these cases, active choice is more important, and individuals should be able to exercise it, even if it leads to a somewhat negative outcome.

Finally, the excessive attention on how nudges affect choices and autonomy has meant that the fact that individual freedom can often result in negative social externalities has not been taken into consideration. Sometimes, what is at stake in terms of public good is more important than individual freedom. As Pérez (2013, p. 3) argues, ‘we should not take individual freedom to be the decisive criterion of differentiation between permissible and impermissible motivational instruments’. Many policy interventions limit freedom of choice. However, even though freedom of choice is compromised, other equally relevant normative objection and policy goals are promoted.

In conclusion, the moral acceptability of nudges changes depending on the issue that is at stake. Influencing choices through non-conscious mechanisms might be problematic in some areas but unproblematic in many others. At the same time, even in domains in which people would like to make

---

<sup>8</sup> Some authors argue that worries about the implications of nudges for autonomy should be balanced with data on their effectiveness. In some domains, even intrusive nudges have moderate effects. Individuals’ preferences over the issue at stake are a crucial factor in predicting the effectiveness of nudges. When individuals have strong preferences for some of the options available, nudges tend to be less effective. For instance, nudges to reduce smoking do not work on smokers, nudges to promote healthy choices tend to be ineffective to change the behaviour of overweight or obese people. These results suggest that nudges actual power to undermine autonomy is often overplayed (Hausman, 2018; Schmidt & Engelen, 2020; Sunstein, 2017a).

their own choices, influencing behaviour and compromising freedom of choice might be justified to promote social goals.

## 5. Concluding remarks

In this chapter, I have discussed the ethical implications of nudges. I have argued that, while the current literature highlights that nudges have an intrinsically problematic normative nature, this is not the case. At the same time, I pointed out that the implementation of nudges raises ethical issues and, accordingly, I have raised some pertinent points to add to the debate on the ethical acceptability of nudges.

Firstly, I suggest that the original normative defence regarding the ethical assessment of nudges should be abandoned. The standard objections to nudges point to essential problems with the original characterisation of nudges. They indicate that the welfare-promoting and freedom-preserving properties of nudges cannot be generally maintained without raising other problematic issues. Objections to the aims of nudges highlight the fact that nudges often fail to improve individuals as judged by the individuals themselves and, when used under this rationale, nudges impose excessive and illegitimate paternalism. Objections to the means of nudges stress that nudges compromise agents' capacity to make their own choices and fail to respect freedom of choice and autonomy. While these objections do not invalidate the use of nudges, they do indicate that their original normative aim is doomed.

Secondly, even though nudges fail to comply with their original normative standards, their use as a policy instrument is not unethical *per se*. I suggest that the ethical implications nudges depend on a combination of three aspects: the aim of the nudge, the type of nudge and the context of use.

In terms of their aims, it is a mistake to assume that the use of nudges is only feasible within libertarian paternalism. Instead, we need to discuss alternative justifications for their implementation. We should explore how to use nudges to promote all kinds of policy standards such as the reduction of inequality, the reduction of negative externalities and the promotion of welfare. In this sense, it is also important to think about how nudges can contribute to better public policy performance by complementing other policy tools, and how nudges can be used to promote policy goals that are already relevant. As policy instruments, nudges should be used to encourage legitimate aims, and their implementation should be subject to standard democratic discussion and control. At the same time, we should be mindful of the fact that nudges are not an all-round solution, and we need to be aware of their pitfalls and limitations.

Thirdly, some objectors worry that nudges work in ways that infringe autonomy beyond the legitimacy of their objectives. To address this objection, I group the central claims into two categories: the rationality objection and the reflective objection. In response to the two complaints, I defend that in both cases, objectors tend to rely on a generalised, misrepresented idea of nudges and on unrealistic interpretations of autonomy not supported by empirical evidence. Reviewing the nature of nudges and

adopting empirically based notions of decision-making highlights the limitations of both objections in the argument that nudges infringe on autonomy.

Fourthly, having a sense of the difference between nudge interventions is crucial when it comes to assessing their ethical implications. Nudges differ in terms of the level of intrusiveness on people's choices; some interventions engage reflection while others reduce reflection and seek an automatic response. To further this discussion, I argue that the principal attempts to distinguish and categorise nudges based on the cognitive mechanism are taking the wrong approach. Cognitive-based distinctions are favoured in previous literature but have significant limitations. Instead, I suggest that the degree of intrusiveness nudges have on people's choices is related more to features about their design and implementation, primarily whether nudges seek an active or passive response, whether they are transparent or non-transparent and the choice situation in which they work (attached or detached from the choice execution context). Nudges that seek a passive response, are non-transparent and not disclosed, and that only work when connected to the choice execution context are more likely to reduce effort, attention and reflection.

Fifthly, while some nudges are more likely to reduce agents' attention over their own choices, I maintain that these nudges are not intrinsically morally problematic. Their implications vary according to the context of choice in which the nudges are applied. I suggest that the ethical implications of nudges depend on the nature of the choices in specific contexts and on the issue at stake. The use of more intrusive nudges might be unproblematic in contexts in which people's choices are already very sensitive to external factors, particularly those coming from private companies, or in contexts in which choices are usually guided by non-reflective decision-making. Likewise, their use is generally unproblematic if nudges target uncontroversial issues. By contrast, their use may be more harmful in decisions in which we are able to engage reflection or in contexts in which the issue at stake is relevant. In general, therefore, it seems that the debate on the ethics of nudges cannot be generalised and should pay more attention to how diverse concerns develop in different situations.

Finally, several questions on the ethics of nudges still remain unanswered. I would like to conclude with a remark on how to advance this debate. The debate on the ethics of nudges should consider people's perceptions and attitudes about the use of nudges. The current discussion is based on abstract notions of freedom of choice, autonomy, manipulation and agency, which, while relevant, do not advance the debate and reinforce the opposition between different philosophical accounts, those in favour and those against nudges. Instead, exploring the acceptance of nudge interventions and complementing the theoretical ideas with folk conceptions of freedom of choice and autonomy could be a useful way to advance the discussion. Public acceptability of nudges helps when it comes to establishing when the use of nudges feels adequate, and how nudge interventions should be implemented and communicated to ensure that they are both, effective and accepted.

## Chapter 3

# The use of behavioural insights in public policy: opportunities and limitations

### 1. Introduction

As explained in the introduction, in the last decade, the use of behavioural economics in policymaking and the implementation of nudges has found widespread acceptance. The incorporation of these factors into policymaking represents a significant development for the formulation of public policies. The evidence from behavioural economics makes it possible to design and implement policies from a more realistic understanding of behaviour, contributes to promoting evidence-based policymaking and broadens the available public policy tools to change behaviour. However, behavioural economics has been criticised for relying on a narrow understanding of rationality and being based on somewhat limited empirical evidence. As a result, critics also underline its limitations as far as informing policymaking is concerned.

Most of the worries about using behavioural economics in policymaking stem from its ties with rational choice theory. As mentioned in chapter 1, behavioural economics first appears as a challenge to rational choice theory. However, it still maintains that this theory is the normative benchmark for behaviour and, therefore, rational choice significantly influences research hypotheses on behavioural economics and how the approach explains empirical findings. The associations between rational choice theory and behavioural economics raise doubts about the main findings on behavioural economics. Most notably, critics argue that the approach overemphasises the prevalence of behavioural biases. As a result, critics also question its value in terms of informing policymaking and argue that behavioural economics might be unhelpful when it comes to identifying reasons for policy interference and nudges might be ineffective at promoting behavioural change (Gigerenzer, 2015, 2018; Grüne-Yanoff & Hertwig, 2016).

Chapter 3 reviews the debate regarding the foundations of behavioural economics and its implications for public policy. The chapter discusses two main themes. Firstly, it acknowledges that previous literature raises essential points regarding the limitations of behavioural economics in terms of informing policymaking. The applications of behavioural economics in policymaking are shaped by how the research programme understands behaviour and are, therefore, influenced by its underlying normative assumptions on rationality. I argue that due to these underlying assumptions, the adoption of behavioural insights in policymaking should be done with caution and hold back from pushing normative claims. Secondly, while reservations about behavioural economics are justified, some of the objections are rooted in theoretical disputes about rationality that overstate the negative implications of behavioural economics. Accordingly, I contend that evidence of behavioural economics is still relevant

in terms of informing policymaking, and nudges have a practical value when it comes to promoting behavioural change.

The chapter proceeds as follows. Section 2 provides an overview of rational choice theory and behavioural economics, and how both approaches influence policymaking. Section 3 analysis the ties between rational choice theory and behavioural economics and introduces the idea of why the association might be problematic for carrying out research on behavioural economics and the policy agenda. Section 4 takes a look at the fast and frugal heuristics approach and its objections to behavioural economics and points out the main problems with the way in which behavioural economics is used in policy contexts. The chapter ends with some concluding remarks on the limitations and challenges of the use of behavioural insights in policymaking.

## **2. Rational choice and behavioural economics: theory and policy implications**

### **2.1. The traditional approach: rational choice theory**

Rational choice theory analyses the way in which agents make decisions that, given specific opportunities and constraints, maximise their utility according to their preferences. While rational choice theory has different interpretations, the neoclassic approach is the most prototypical and influential. This approach models behaviour under specific assumptions in terms of agents' preferences, beliefs and computational abilities (Hedström & Ylikoski, 2014).

Firstly, the theory assumes that agents have consistent preferences that comply with the following axioms:

- i. Completeness: agents are always able to rank alternatives; they can express a stronger preference for one or the other in any pair of options.
- ii. Transitivity: the order of different alternatives prevails. If an agent prefers A to B and B to C, they will prefer A to C.
- iii. Context independence: for two complete states of the world, preferences between options never depend on what other alternatives are available.
- iv. Choice determination: preferences imply determining choices. If the highest-ranked preference is feasible, then the agent chooses the highest-ranked option.
- v. Time consistency: preferences are supposed to be stable in time. If an agent prefers A to B at  $t_0$ , they will also prefer A to B at time  $t_1$ .

When agents' preferences meet these requirements, they are said to be consistent and can be represented with a utility function that is used to explain and predict people's choices. The canonical rational choice theory does not judge the substantive content of preferences; however, models of rational choice often assume that self-interest is the primary motivation for action.

Secondly, the theory assumes that agents have rational beliefs and unlimited cognitive capacities. Rational choice expects agents to have complete and correct information about everything that is relevant to a particular context. Likewise, it is implicitly assumed that agents can assess the possible consequences of different actions and correctly calculate their probability and expected utility, allowing them to then choose the option that maximises it (Hedström & Ylikoski, 2014).

These assumptions about agents' preferences, beliefs and computational abilities are at the core of the *expected utility model* and the *discounted utility model*, two of the most important models in neoclassical economics. The *expected utility model* analyses the way in which agents make choices in situations of risk and uncertainty. The model states that agents are expected to choose the option with the highest expected utility when presented with a number of options. The expected utility of an action is determined by the utility of each of the possible outcomes, weighted by the probability of each outcome conditional on the actions of the agent (Briggs, 2019). The expected utility model attributes agents with greater deliberative and computational abilities. It assumes that agents assess the situation, consider the consequences of their different actions, calculate their probability of occurring and their expected utility and accordingly choose the optimising option<sup>9</sup>. The discounted utility model is used to study intertemporal choice and accounts for the phenomena of temporal discounting, which describes agents' general preferences for small but present rewards over delayed, greater rewards (Samuelson, 1937). The discounted utility model establishes that agents behave to maximise the sum of future utilities, appropriately discounted by the discount rate. The model assumes that discounting is exponential and has a constant discount rate; it also assumes that the utility of an outcome is independent of previous consumption and remains constant across time (Elster, 2000; Loewenstein & Elster, 1992).

Within rational choice theory, the term *homo economicus* is employed to describe the model of human behaviour used to analyse individuals' decisions. *Homo economicus* is rational, has consistent preferences, unlimited cognitive abilities, and seeks to maximise utility. The model describes an agent who considers all of the relevant information when facing a decision, evaluates the probability of each option and its potential consequences, and chooses the option that maximises self-interest based on an accurate cost-benefit analysis.

The *Homo economicus* model, the expected utility model and the discounted utility model are used to study economic decisions, but also extensively used to study behaviour in non-economic fields. Likewise, rational choice theory, and the *homo economicus* model in particular, have a defining role in

---

<sup>9</sup> Illustrated with the classic umbrella case, the utility of bringing an umbrella on a long walk depends on agents' preferences and how their actions drive their preferences. Peter prefers to leave the umbrella at home so that he is more comfortable on the walk; however, he does not want to get wet if it rains. The situation has two possible states, either it will rain, or it will not. Peter can act in two ways, take the umbrella or leave the umbrella. There are four possible outcomes: it rains, and he has taken the umbrella, it rains, and he has not taken the umbrella, it does not rain, and he has taken the umbrella, and it does not rain, and he has not taken the umbrella. The expected utility of taking or leaving the umbrella depends on the utility (value) assigned to each of these outcomes, weighted by the value that Peter's actions intend to deliver (Briggs, 2019).

policymaking. While the model's influence is perhaps not explicit, it outlines the way in which agents are expected to behave and the solutions and strategies available to change behaviour.

Traditional approaches to policymaking are based on the assumption that individuals are rational and act in a self-interested manner. The model states that the rational strategy that benefits self-interest is non-compliance in many situations at stake in policymaking. For instance, as mentioned in chapter 2, agents are expected to behave as free riders in situations concerning the distribution of non-excludable and non-rivalrous goods: i.e., not contributing to the production of these goods but still taking advantage of their benefits. Overall, the assumption is that rational interest favours non-compliance in many situations addressed by policies, and that agents have no intrinsic motivations to comply in the absence of deterrent measures (Pettit, 1996).

In policymaking, the approach prescribes the use of monetary incentives, sanctions, and the provision of information as the main strategies for influencing behaviour. Traditional institutional design tools are designed to apply to self-interest. The most relevant policy tool is the use of monetary incentives. Positive incentives, such as subsidies, and mostly negative incentives, such as taxes, are used to modify decisions across domains. Incentives affect the behaviour of agents as a result of their self-interest, thereby altering the cost-benefit of the options to make them more or less attractive. The second most extended policy tool is the use of filters or selection mechanisms. Positive filters such as licences and negative filters such as bans, sanctions and vetoes select types of agents or options. They restrict agents' opportunities by making options available or unavailable, and making the unavailable less attractive by increasing their cost with sanctions and fines (Pettit, 1996). Finally, the last available institutional design tool is the provision of information. Rational choice theory suggests that when agents act against their self-interest, even when incentives and selection mechanisms are in place, it is because they lack information. Accordingly, providing information helps individuals to re-adjust their preferences and behaviour to the specific opportunities and constraints that they face.

Overall, the lessons of rational choice for institutional design are straightforward. Self-interest favours non-compliance, and in the absence of these mechanisms, it is risky to expect agents to be motivated to comply. Incentives and selection mechanisms work because they appeal to self-interest. The provision of information works because it allows agents to assess the options and select the optimum choice.

The standard rational choice theory approach and the homo economicus model have been highly influential in theory and policy. Rational choice is the central normative theory in economics, and it is still used to describe behaviour. Likewise, it remains the most important unifying theory on decision making and has had a significant influence on other disciplines, such as sociology and political sciences. As noted, its relevance is also patent in policymaking, in which the assumptions of self-interest and rationality underlie the main strategies used to change behaviour.

However, as has been proven, extensive evidence contradicts rational choice theory. Numerous experimental findings, both in the lab and in the field, challenge its predictive and descriptive power. The models in rational choice research, the expected utility model and the discounted utility model have proven unsatisfactory when it comes to describing behaviour. Likewise, extensive empirical research also contradicts the assumptions of preference consistency, unbounded rationality, unbounded willpower and unbounded selfishness (Mullainathan & Thaler, 2000). The evidence suggests that rational choice theory highlights important behavioural motivations and might work for analysing behaviour in situations of certainty with specific conditions and limited alternatives. However, it is unquestionable that it has significant limitations when it comes to describing human decision-making in most real situations.

## **2.2. Behavioural economics: new insights and new policy implications**

Behavioural economics emerges as a challenge to the explanatory power of rational choice theory. Research in the field provides empirical evidence that agents' behaviour systematically departs from rational decision-making and accordingly attempts to increase the descriptive and predictive power of economic models by offering more realistic social and cognitive foundations (Angner & Loewenstein, 2007). The contributions to research on behavioural economics are broad and diverse. However, as explained in chapter 2, insights from behavioural economics can be broadly classified into three main groups:

- (i) contributions regarding bounded rationality (referred to in *'Nudge'* as "heuristics and biases"),
- (ii) contributions regarding time discounting (referred to in *'Nudge'* as "temptation") and
- (iii) contributions regarding the effect of social preferences and social influence (referred to in *'Nudge'* as "following the herd").

The contributions under the bounded rationality category challenge expected utility theory and the assumptions of consistent preferences and perfect cognitive abilities. Extensive evidence from research on judgment and decision-making shows that subjects struggle to assess probabilities and process information and that their decisions are affected by heuristics and biases. As far as judgments of probability under risk and uncertainty are concerned, early work done by Kahneman and Tversky (1979, 1981) showed that people's assessments of utility depend on reference points and are sensitive to whether options are presented as losses or gains. These findings indicate that people's preferences violate the axiom of transitivity, which assumes that preferences between options should not vary according to how options are presented. Behavioural economics has developed an alternative descriptive model of choice under risk: *prospect theory*. Some of the most important biases identified by this research include:

- framing effects, which occur when people's preferences depend on how the options are presented to them,

- loss aversion, which describes people's tendency to prefer to avoid losses rather than obtain equivalent gains,
- the endowment effect, which shows that people's preferences are not independent of their endowment and tend to assign more value to what they have than to what they do not have,
- the status quo bias (or inertia), which describes the preference for the actual state of affairs, and
- risk-seeking and risk-averse preferences, which specify that agents are risk-averse towards positive outcomes and risk-seeking towards negative outcomes

As far as general judgments and information processing are concerned, behavioural economics has also documented the effect of systematic heuristics and biases (Kahneman et al., 1982). The heuristics and biases programme identifies specific mechanisms underlying people's judgment. Some of the most important heuristics and associated biases include:

- the availability heuristic, which refers to the tendency to judge the probability of an event occurring by how quickly (how available) examples or instances of the same case come to mind,
- the representativeness heuristic, which accounts for the tendency to overestimate the probability of an event, by judging it according to how representative of a particular category it is and not by how likely it is to occur,
- the anchoring and adjustment heuristic that describes the effect that irrelevant available information (the anchor) has on people's decisions,
- the optimism bias, which describes people's tendency to overestimate the probability of positive events and underestimate the probability of adverse events, and
- the saliency bias, by which agent's judgments about factors, people and situations are not driven by what is relevant but by what is most noticeable.

The findings of the heuristics and biases programme contradict the assumptions of rational choice theory because they highlight the fact that agents struggle to process information. This directly challenges rational choice because it violates the axioms of the expected utility model. Given agents' cognitive limitations, heuristics help people to navigate uncertainty and make decisions in situations of complexity or information overload; however, heuristics tend to lead to mistakes and biases (Kahneman et al., 1982; Tversky & Kahneman, 1974).

The contributions under the time discounting category study choices between options whose consequences occur at different points in time. Empirical research identifies time-inconsistent preferences and problems of self-control and directly challenges the discounted utility model (Loewenstein & Elster, 1992). Research on temporal choice in behavioural economics documents the existence of non-exponential discount rates which cause time-inconsistent preferences. Empirical research also documents intra-individual differences of discount rates depending on the domain (Loewe,

2006). A recurrent finding regarding choice over time is that agents that are faced with a choice between an inferior early option and a superior later option tend to prefer the latter when both options are remote but switch to former as both approach in time (Elster, 2000). Intemporal preferences explain problems of self-control. When agents evaluate the utility of outcomes in a distant future, they are more rational and plan to behave in ways that maximise their long-term interest. However, when choices approach in time, they change their preferences and prefer immediate benefits (DellaVigna, 2009). Behavioural economics uses the *hyperbolic discounting model* to study time-inconsistent preferences and their implications (Laibson, 1997). Some of the more significant effects identified under hyperbolic discounting model include:

- myopic behaviour, which describes how people tend to focus on the present and short-term options and ignore future consequences, and
- lack of self-control, which explains agents' problems of sticking to actions that benefit their long-term wellbeing.

Finally, contributions under the social preferences and social influence category report findings that emphasise the fact that agents' behaviour is not always motivated by self-interests, does not occur in isolation and is, therefore, affected by the behaviour of other agents<sup>10</sup>. In rational choice, individual self-interest is often considered the primary motivation for action; however, experimental evidence shows that agents do not always seek an outcome that maximises their benefits. Some of the most significant mechanisms identified in this regard include:

- fairness, which refers to agents' preferences for equal treatment and accounts for inequality aversion,
- altruism, which references preferences to promote the wellbeing of others unconditionally, even assuming costs or losses for oneself,
- moral norms: which are non-instrumental mandates to act or refrain from doing so that are followed unconditionally (Elster, 2007),
- reciprocity, which references the preferences for cooperative behaviour in response to other cooperative behaviour.

Research on social influence analyses the effects of preferences beyond self-interest and the effect of interpersonal influences and tendency of agents to conform to the behaviour of others. Some of the most significant mechanisms identified within research on social influence include:

---

<sup>10</sup>The evidence that people have limited self-interest and that they are also driven by social preferences and affected by social influence does not belong exclusively to behavioural economics. It has been advanced by game theory research and by approaches to rational choice that use a broader interpretation of the theory, especially with contributions from sociology (Hedström & Ylikoski, 2014).

- social norms, which refer to regularities of behaviour that agents follow because they believe that most people in their reference group follow them and because they believe that most people in their reference group believe they should be followed. Social norms are seen as socially “appropriate” or “correct”, and violating them might trigger negative emotions, such as shame or guilt (Bicchieri, 2005, 2016; Elster, 2007),
- social conformity, which describes the spontaneous convergence with observed or perceived average social behaviour,
- rational imitation, which describes how the behaviour of others modifies agents’ beliefs in contexts of uncertainty, because it provides information about what might be the rational course of action, and
- strategic interaction, which indicates that an agent’s cooperation in promoting a social good, for instance, is conditional on the cooperation of others (Bicchieri, 2005; Elster, 2007).

All of these factors challenge some of the rational choice theory assumptions because they stress that agents’ behaviour is not always self-centred or egoist.

Overall, behavioural economics provides systematic evidence that agents exhibit limited self-interest, limited will power, preference inconsistency due to time and situation, and difficulties in processing information due to cognitive constraints (DellaVigna, 2009). Based on these findings, research on behavioural economics emphasises that agents exhibit systematic and predictable rationality failures, and they often behave in ways that undermine their own wellbeing. The term *internality* is used to reference the cost in subjective wellbeing resulting from failing to behave in ways that maximise long-term interest.

In light of the findings from behavioural economics, the assumption underlying traditional policy design strategies appear limited and insufficiently based on empirical evidence. As a result, proponents of the approach emphasised the importance of incorporating insights from behavioural economics into policymaking to design and implement policy interventions grounded in a more realistic understanding of behaviour and, thereby, improve their effectiveness and efficiency. As mentioned in Chapter 1, the publication of ‘*Nudge*’ was a crucial contribution in this regard and, since then, the approach has solidified its relevance in policymaking.

Behavioural economics informs policymaking in three main ways. Firstly, the approach is praised because it provides a more realistic account of behaviour that improves the insights used to design and implement policies and helps to get a better understanding of how existing policies will perform. Secondly, the approach uncovers new reasons for policy intervention and stresses the need for interfering in areas in which individuals might have problems behaving in ways that benefit their wellbeing. Along the same lines, behavioural economics emphasises the need to correct internalities and favours interventions that have a positive effect on people’s subjective wellbeing. Finally, the

approach develops new tools for behavioural change: nudges. Insights from behavioural economics make it possible to design interventions that make options more attractive or less so in social and cognitive terms (Bhargava & Loewenstein, 2015; Chetty, 2015; Thaler & Sunstein, 2008).

The distance between rational choice theory and behavioural economics seems to be clear. Theoretically, the two approaches employ different models to describe and analyse human decisions. Their policy implications also differ, and the theories prescribe different strategies and policy tools for behavioural change. The findings from behavioural economics challenge rational choice and uncover its descriptive limitations. Nonetheless, behavioural economics studies deviations from rational choice theory and, therefore, rational choice is still relevant for the purposes of research on behavioural economics. The status of rational choice within behavioural economics raises doubts about the reliability of findings from behavioural economics and their implications for policymaking.

### **3. The ties between behavioural economics and rational choice and their implications**

Initially, findings from behavioural economics were received with scepticism within rational choice theory. Advocates of rational choice contended that the factors identified in research on behavioural economics were irrelevant because market competition should eliminate or mitigate their effects, because agents would act rationally when facing higher incentives and because subjects were able to overcome biases through learning (Mullainathan & Thaler, 2000). However, repeated empirical analysis depicting the descriptive inadequacies of rational choice prove that these factors were unignorable. Rationality deviations were found to be systematic, predictable and challenging to overcome. As a result, behavioural economics is gaining recognition within economic research.

As explained in the previous section, behavioural economists analyse individual decisions from a more realistic understanding of behaviour and accordingly develop models that adjust or replace rational choice theory models. However, the integration on behavioural economics into economic research has been favoured because behavioural economics remains committed to rational choice. Rational choice theory is a normative and descriptive theory and, as such, it can be looked at in both ways. It is a normative theory because it defines how behaviour ought to be and how agents are expected to behave to maximise their utility. It is also a descriptive theory because models of rational choice were and are still used to study agent's behaviour in several situations (Angner, 2016). Behavioural economics represents a development because it descriptively challenges rational choice; however, it does not challenge its normative assumptions. Rational choice models are used in behavioural economics as 'implicit and explicit benchmarks against which many of the authors contrast their contributions' (Loewenstein & Elster, 1992, p. 9).

Behavioural economists note the affinity between rational choice theory and behavioural economics. For instance, Angner & Loewenstein, (2007) consider that behavioural economics emerged as a sub-discipline within economic research, and Chetty (2015, p. 1) stresses that 'behavioral economics

represents a natural progression of (rather than a challenge to) neoclassical economic methods'. The consensus is, therefore, that rational choice theory is the normative theory of behaviour, whereas research on behavioural economics is used to analyse how agents actually behave in real situations.

The influence that rational choice has on research on behavioural economics is substantial. As the normative understanding of behaviour, rational choice influences the development of specific research hypotheses and guides the interpretations of empirical findings. The association is essential to understanding the main claims of behavioural economics. It is by using rational choice theory as a benchmark that it is possible to identify behaviours and decisions that do not conform to what is expected and accordingly interpret them as biases. In that sense, the idea that systematic flaws in human decision-making affect behaviour implicitly accepts rational choice theory as the normative standard of decision-making and, consequently, considers consistent deviations for this standard as systematic and predictable irrational biases.

As argued in Chapter 1, the association is also crucial when it comes to supporting behavioural recommendations in the policy context. Firstly, the empirical findings of behavioural economics support the idea that making options more or less attractive in cognitive and social terms works as far as changing agents' behaviour is concerned and, therefore, that nudges can influence people's choices. Likewise, the notion that these factors affect behaviour systematically and predictably is supported by the interpretation of empirical findings that behavioural economics makes based on the assumptions within rational choice. Finally, the influence of rational choice is more direct on the reasons for interference defined by behavioural economics. The notion that biases undermine people's wellbeing is crucial when it comes to supporting the recommendations of behavioural economics on policy interference in areas in which people apparently suffer from the effects of irrational biases. This argument serves to justify nudging as a strategy that ought to prevent people from engaging in poor choices and help them choose what is best for them (Thaler & Sunstein, 2008)

Whether the normative status of rational choice in research on behavioural economics is positive or negative is up for debate. Some authors argue that using a theoretical background and specific assumptions contributes to the progression of theory. Guala (2000) argues that normative theory is relevant for guiding the testing and interpretations of empirical findings and also makes it possible to reformulate normative models. Similarly, Angner (2015) argues that neoclassical theory within economics serves as an *ideal type* that is useful as a reference or as an analytical tool to guide empirical research even when it is rejected by descriptive findings. While perhaps not explicit, this seems to be the agreement in research on behavioural economics. Nonetheless, other contributions emphasise that the epistemic status of rational choice in behavioural economics leads to methodological problems, has negative implications for theory development, and might be specifically problematic as far as informing public policy is concerned (Elqayam & Evans, 2011; Gigerenzer, 2015, 2018; Małecka, 2020).

At a theoretical level, the association can be problematic. Elqayam & Evans (2011) discuss the methodological problems of employing normative theories in research on judgment and decision-making. The authors stress that *normativism*, understood as the commitment to an idea of how decision-making ought to be, affects what is studied, how it is studied and how the findings are interpreted and, therefore, leads towards incorrect assessments of empirical findings<sup>11</sup>. The main problem is that normative theories cannot be tested empirically because empirical facts are interpreted in the light of the normative agenda. Normativism does not contribute to advancing research and knowledge on decision-making because there are competing normative accounts of judgment and decision-making. The authors underline the fact that normative theories promote research on how people do not behave rather than how they actually do behave, which is what should be relevant in the first place. Their criticism is not expressly aimed at the heuristics and biases programme (the approach to research on judgment and decision-making in behavioural economics), although some of their concerns apply. Behavioural economics is centred on testing deviations from rational choice and defines specific decision-making mechanisms as biases because they do not conform to what rational choice predicts. Therefore, it seems that using rational choice as a normative backdrop might be problematic.

In a more specific criticism of behavioural economics, Małecka (2020) also argues that the normative status of the expected utility model has been widely accepted in economics. However, its implications have not been properly taken into consideration. Małecka (2020) contends that it is unclear why economics needs a normative theory and a descriptive theory and stresses that moving beyond expected utility, which has already been proved inaccurate when it comes to describing behaviour, might be a useful way to advance research.

However, as far as theoretical problems in behavioural economics and their negative implications for policymaking are concerned, the most confrontational criticism comes from the *fast and frugal heuristics approach*. The following section takes a look at this objection and what it implies for the use of behavioural insights in policymaking.

#### **4. Objections to behavioural economics and its use in policymaking: the critic from the fast and frugal heuristics approach**

The fast and frugal heuristic research programme explores the cognitive mechanism underlying judgment and decision-making, taking bounded rationality into account. The approach argues that, in everyday decision-making, agents frequently use fast and frugal heuristics to make decisions. Heuristics are simple rules of thumb that are quick and cognitively cheap and that in contexts of complexity, uncertainty and time pressure allow agents to come up with satisfactory decisions (Gigerenzer, 2008). The fast and frugal heuristics programme studies decision-making by taking the relation between

---

<sup>11</sup>The problem is not with formal theories or models, which can be tested empirically. The problem is with the normative defence of a particular notion of rationality and decision-making.

agents' cognitive abilities and the structure of the environment into account. The essential idea is that agents have limited cognitive abilities, and they often have to make decisions in complex and uncertain environments. As a result, agents have developed a series of adaptative heuristics that, given their bounded rationality, help them to quickly navigate information, learn how to stop the information search and utilise this information to come up with good decisions. The research programme has two main components. Firstly, a study of "the adaptive toolbox", which refers to the identification of heuristics that work well in different environments, and a study of the structure of these heuristics, referred to as their "building blocks". Secondly, the study of "ecological rationality", which focus on discovering the environments in which heuristics work and the contexts in which they fail, in order to understand how to improve decision-making by teaching heuristics or by re-thinking the structure of the environment (Gigerenzer, 2008; Gigerenzer & Todd, 1999).

Despite studying judgment and decision-making and building on the notion of bounded rationality, the perspective on the fast and frugal heuristics programme contrasts with the contributions about bounded rationality defined in the heuristics and biases programme. The essential difference is the positive or negative outlook towards heuristics and bounded rationality. As mentioned in the previous section, the heuristics and biases research programme portrays heuristics as problematic and emphasises that agents' decision-making is flawed and affected by systematic and unavoidable biases. By contrast, the fast and frugal heuristics programme argues that heuristics are smart and allow individuals to resolve real problems in contexts of uncertainty.

The fast and frugal heuristics approach builds on Herbert Simon's notion of bounded rationality. Simon was one of the earliest critics of rational choice (Nagatsu, 2015a). The author developed a critique of neoclassical economics that was explicitly critical about the behavioural assumptions underlying economic models. Simon argued that neoclassical economics relied on simple assumptions about behaviour, lacked empirically grounded theories about how economic actors actually make decisions in the face of uncertainty, and, as a theory, it was unable to explain actual human behaviour in most real-life situations. Simon distinguished between *substantive* and *procedural* rationality. According to Simon, neoclassical economics is a theory of substantive rationality; it serves as a theory to predict choices, that takes into account external constraints and expects agents to adapt to the situation to maximise their utility. However, the theory overlooks the actual decision-making process. Rational choice theory analyses individual behaviour by employing assumptions about agents' internal decision-making process (as explained, perfect rationality, perfect foresight, stable preferences and unlimited computational abilities) and describes and predicts behaviour *as if* subjects actually behave as stated. By contrast, Simon argues the necessity to study procedural rationality, but rather than focusing on the results of decisions, he emphasised the need to study agents' decision-making processes (Simon, 1959, 1976).

Simon originally coined the concept of “bounded rationality” to advocate for an alternative approach to studying decision-making that took into consideration the objective environment and the perceptual and cognitive processes of actors. According to Simon, in order to study decision-making, specifically decision-making under uncertainty, it is mandatory to take the environmental constraints into consideration, and also study and explain how agents understand and perceive the environment and act upon it. Simon also questioned the notion of rationality as optimising behaviour. His research pointed out that optimising is either impossible or too costly (requires too much cognitive effort) in real-life choices. He developed the notion of *satisficing* as an alternative, which captures the idea that, when considering agents’ cognitive abilities and the characteristics of the environment, individuals usually search for a *good enough* answer rather than optimising one (Simon, 1997)<sup>12</sup>.

The fast and frugal heuristics approach follows Simon’s behavioural critique of neoclassical economics and rejects the notions of perfect rationality and the idea that the ultimate goal is the need to optimise. In contrast with heuristics and biases research, the fast and frugal heuristic approach is fundamentally a descriptive theory, not committed to a normative idea of rationality and one that does not attempt to test whether agents’ decision-making follows specific normative notions of behaviour<sup>13</sup>. As a result, the approach is critical of the normative standards underlying behavioural economics and the heuristics and biases research programme. Gigerenzer (2018, 2015) emphasises that while heuristics and biases research started as an attempt to increase the psychological foundations of economic models, by remaining committed to rational choice theory it has turned into a study of deviations of rational choice, which has problematic consequences for the development of research and the development of policy recommendations.

The disagreements between the heuristics and biases approach and behavioural economics versus the fast and frugal heuristics approach are well-grounded. Some authors emphasise that the two approaches are contradictory and in dispute (Grüne-Yanoff & Hertwig, 2016; Katsikopoulos, 2014; Samuels et al., 2012). Proponents of the fast and frugal heuristics programme tend to be very critical of the findings on behavioural economics. The essential objection is that behavioural economics overstates the notion that individuals suffer from systematic and predictable behavioural biases. Gigerenzer (2018, p. 307) explicitly states that behavioural economics suffers from *bias*: ‘the tendency to see systematic biases in behavior even when there is only unsystematic error or no verifiable error at all’. He also points out that the evidence for biases mainly comes from experimental findings that lack external validity.

---

<sup>12</sup>Nagatsu (2015a) and Sent (2004) refer to Simon’s contributions as “old behavioural economics”, an approach that is an upfront criticism of both the descriptive and normative dimensions of rational choice theory. By contrast, they refer to the heuristics and biases programme and the research on hyperbolic discounting as “new behavioural economics”, contributions that challenge the descriptive dimension of rational choice but not its normative assumptions. Simon’s contributions were never as influential in economics as the contributions from “new behavioural economics” because they were more critical of the standard approach.

<sup>13</sup>Please note that this is why the fast and frugal heuristic research has been criticised for providing *ad hoc* explanations and only focusing on cases in which heuristics work well.

Accordingly, he argues that the picture of rationality portrayed in behavioural economics is incorrect and emphasises three main points. Firstly, he argues that behavioural economics lacks evidence to support the premise that agents are predictably irrational, and their decisions are systematically affected by cognitive biases. Secondly, he argues that behavioural economics lacks evidence that these biases cannot be corrected or overcome with education. Finally, he argues that it also lacks evidence that these biases have substantial welfare costs.

In response to the evidence from behavioural economics, fast and frugal heuristic research suggests that factors that behavioural economics depicts as biases can actually be conceived as “ecologically rational”, i.e., rational in the context in which agents make decisions. As argued in Chapter 2, the idea that framing effects are irrational is based on the argument that the logical information presented to the subjects is identical. However, Gigerenzer points out that the definitions are not totally equivalent; instead, he suggests that frames provide extra information about the situation and that heuristics can recognise this information and adapt the response to it. For instance, when a medical professional frames the chances of survival as positive (90% chance of survival) or negative (10% chance of death), agents perceive the positive or negative frame and adapt their decision accordingly. This is “ecologically rational” because, in a context of uncertainty, it provides additional clues about what might be the best outcome. Another example covered by fast and frugal research is the premise that problems with assigning probabilities tend to be diminished when they are presented with natural frequencies, rather than with probabilities. In these cases, changing the environments so that they match our adaptive heuristics reduces errors in judgment (Gigerenzer, 2015, 2018).

Overall, fast and frugal research argues that people are “ecologically rational”, emphasises that biases are less prevalent than portrayed in research on behavioural economics and suggests that these errors can be avoided in contexts in which mistakes happen by teaching people specific heuristics or decision-making mechanisms and/or by accommodating specific elements from their environment. As argued, these insights pose problems for the way in which behavioural economics informs policymaking in two main ways. Firstly, if biases are not frequent and have been overstated, nudges might be ineffective when it comes to changing behaviour. Secondly, and more directly, if decision-making is not biased and does not generally impact on agents’ wellbeing, nudging to promote agents’ wellbeing is undesirable.

Recently, and as a response to the use of nudges, the fast and frugal heuristics research programme conceptualised how their empirical findings could be translated to policymaking. Building on their theoretical findings, the research programme endorsed the use of what they called “boost policies” (Grüne-Yanoff et al., 2018; Grüne-Yanoff & Hertwig, 2016; Hertwig, 2017; Hertwig & Grüne-Yanoff, 2017). I examined what they are and why they are different from nudges in the following section.

#### 4.1. Boost policies: characteristics and justification

As nudges, boosts are non-regulatory, non-economic interventions for behavioural change informed by research on bounded rationality. However, instead of being designed to steer people's choices towards a specific outcome, boosts are interventions that try to improve individuals' decision-making abilities or competence in making good choices. Examples of boost interventions include:

- i. interventions that target individual skills and knowledge, for instance, educational programmes specifically inspired by behavioural insights,
- ii. interventions that target the available set of decision tools by, for instance, teaching heuristic decision-making, and
- iii. interventions that target the environments in which decisions are made, for instance, changing or adapting the structure of the information or the available options.

Some examples of boost policies in practice include the introduction of probability and statistics courses in schools to improve agents' capacity to understand information on probability; the teaching of decision trees and heuristics to professionals in different fields to improve their ability to identify problematic cases and solve them; the use of courses to teach financial literacy, and the provision of information<sup>14</sup> intended to improve people's competences, for instance, information that presents the symptoms of a stroke so that, if it happens, agents can recognise it and act accordingly (Grüne-Yanoff & Hertwig, 2016; Hertwig, 2017).

In building the argument in favour of implementing boost policies, proponents of boosts have tended to emphasise the ways in which boosts are better than nudges. Grüne-Yanoff & Hertwig (2016) argue that boosts influence behaviour openly and favour reflective thinking. They also remark that, unlike nudges, boosts do not assume or are required to know an agent's preferences and goals. Likewise, they highlight that boosts designed to promote malevolent purposes are easier to detect than "evil nudges". Finally, they explicitly emphasise that because boosts require agents to have some degree of motivation and competence, 'the criticism that nudge policies infringe on human autonomy and dignity does not apply (or applies less) to boost policies' (Grüne-Yanoff & Hertwig, 2016, p. 176).

Another difference is the underlying purpose of boost policies. The main inspiration behind boosts is the idea that behavioural biases can be overcome by learning and improving agents' decision-making skills. Proponents of boosts highlight the fact that nudging assumes that people are doomed and cannot improve their cognitive abilities. Boosts, instead, have an explicit commitment to improving decision-making and empowering people to make good choices. Accordingly, proponents of boosts contend that boost policies might be more effective than nudges when used to promote long-term behavioural change because the effects of boosts are more likely to persist when the interventions are removed than the

---

<sup>14</sup> Provision of information not intended to persuade agents to do something.

effects of nudges. As a result, boosts are not only defined as being well-intentioned, but are also defended as being more effective at overcoming the effect of biases (Hertwig & Grüne-Yanoff, 2017; Hertwig & Ryall, 2016).

Overall, fast and frugal heuristics research emphasises that insight on human behaviour provides inspiration for implementing boost policies aimed at improving people's competences and their ability to make good choices (Hertwig, 2017). Likewise, to defend this approach, proponents of boosts present them as being less vulnerable to the normative objections that nudges face and as a more ethically permissible behavioural change intervention (Sims & Müller, 2019). While the boost versus nudge rhetoric has been toned down in more recent publications (e.g., Hertwig, 2017), its contraposition remains central to the discussion and their respective translations to public policy have entrenched rhetorical divisions further.

#### **4.2. Is there a real dispute? Core claims vs rhetorical claims**

As previously mentioned, literature exploring the debate between the two approaches to bounded rationality usually presents them as contradictory and even irreconcilable views on human decision-making (Grüne-Yanoff & Hertwig, 2016; Katsikopoulos, 2014). Research in this area has been salient in pointing out areas of disagreement, and the argumentation on both sides has sometimes intensified this debate (e.g., Gigerenzer (1991), Tversky & Kahneman (1996) and Lewis (2016)). As explained, the recommendations of both programmes for policymaking appear to have heightened disagreement by discussing which of the programmes should inform policymaking and which interventions, nudges or boosts are more effective and more ethically permissible for behavioural change.

However, Samuels et al. (2012) argue that the dispute between the heuristics and biases approach and the fast and frugal approach has been overstated and that the empirical findings and interpretations of reasoning and decision-making for both research programmes do not translate into a direct conflict between the two. The authors stress that the dispute between the two approaches is a matter of “rhetorical excesses” and that while there are some major disagreements, there is also common ground. To support their argumentations, Samuels et al. (2012) distinguish between “core claims” and “rhetorical flourish”. They use the term “core claims” to refer to a claim that is central to the research programme that is supported by empirical evidence and that advocates of the programme are prepared to endorse in their more careful moments. By contrast, they used the term “rhetorical flourish” to refer to arguments that are not central to the research programme, not fully supported by empirical evidence and that proponents of the programme do not typically endorse when being careful and reflective.

According to the distinction, the authors stress that the approaches offer compatible conclusions on how agents deal with judgment and probability. As explained, the heuristics and biases approach claims that people's probabilistic judgments deviate from the norms of rationality prescribed by rational choice theory. This is a “core claim”, supported by empirical evidence. This claim, however, has been

embellished by “rhetoric excesses” when proponents of the approach argue that people are systematically irrational, and that heuristics and biases are impossible to overcome and always lead to negative outcomes. The fast and frugal heuristic approach stresses that people have heuristics that allow them to correctly assess probabilities when problems are reformulated in terms of frequencies. This is also a “core claim”, supported by empirical evidence. However, by accepting this claim, the approach also recognises that, in some contexts, adaptive heuristics are not available or not ecologically rational. As a result, the claim that heuristics are always ecologically rational and always lead to good answers has also been overstated and is not supported by empirical evidence. According to the distinction, Samuels et al. (2012) emphasise that ‘neither research program denies the core claims of the other, and in many cases, it is clear that they should endorse each other’s core claims’ (Samuels et al., 2012, p. 36).

Samuels et al. (2012) also suggest that normativism has affected how the two programmes approach research and interpret empirical findings. The authors suggest that heuristics and biases research is concerned with finding cases where heuristics fail and lead to biases. By contrast, fast and frugal heuristics research is concerned with finding cases in which heuristics do a good job. They also emphasise that the approaches have tended to develop research in non-intersecting domains and with limited acknowledgement of the findings of the opposite approach. As a result, the development of both approaches might reinforce their “rhetorical claims” instead of exploring areas of agreement.

Finally, the conflict between the two approaches appears overstated because the programmes cannot provide conclusive information about judgment and decision-making. The mechanisms underlying judgment and decision-making are still unclear, but what emerges from the two approaches is both relevant but also disputable. In fact, recent publications identify problems with the arguments and findings of both approaches (e.g., Dhimi et al., 2019; Gal & Rucker, 2018). This suggests that, while both programmes provide interesting evidence, their results should not be understood as truisms. At the same time, none of the findings from one approach totally invalidate the findings from the other.

Taking the conflict between the two approaches at a theoretical level into consideration, their translations into public policy should be very different. However, the main discrepancies are not with the way in which boosts and nudges work, their potential uses and their potential effectiveness. Instead, the objections emphasise the negative features of nudges arising from their underlying normative assumptions (Sims & Müller, 2019). As a result, the policy agendas of both approaches make it clear that the dispute is overstated and mostly results from disagreements about the normative ideas on rationality. Oddly, Grüne-Yanoff & Hertwig (2016) even argue that nudges and boosts might propose identical policies. However, the interpretation of which causal mechanism they trigger differs and, therefore, their implications for policymaking are also different. As regards the use of default settings Grüne-Yanoff & Hertwig (2016) argue:

'it might seem that nudge and boost policies share the same means. Yet a more detailed analysis shows that this is not the case. The SH and H&B programs tend to disagree about the underlying causal mechanism that explains the relationship between default setting and changes in choice distribution. H&B researchers tend to explain default effects in terms of "inertia, status-quo bias, or the 'yeah, whatever' heuristic" (Thaler and Sunstein 2008, p. 83). This kind of explanation stresses the biasing features of setting any default, thus revealing the policy to be re-biasing. SH researchers, in contrast, explain default effects in terms of the implicit recommendation or endorsement effect (e.g., Gigerenzer and Brighton 2009, p. 130; see McKenzie et al. 2006). This kind of explanation stresses the genuine social information contained in the default, thus describing the behavioral change in response to the default as consisting in a learning effect, and hence revealing the policy to be debiasing. Thus, even in cases where nudgers and boosters propose the same policy, their respective mechanistic interpretation of the intervention distinguishes the distinct goals they pursue with it'. (p. 163).

Recently, proponents of boosts have been presenting examples of boosts that clearly differ from nudges, and characterise boosts as policies that require the motivation of the targeted agents and engage the competences of these agents in some way (Hertwig, 2017; Sims & Müller, 2019). However, the quote illustrates that the emphasis is not on how the strategies differ in the way they affect behaviour but on their normative appeal. Boosts are preferred, not because they work better than nudges, but because they build on a more admirable idea of human behaviour.

Borrowing a term from Samuels et al. (2012), I suggest that the dispute between nudges and boosts and their implications for policymaking are also a matter of "rhetorical excess". The "rhetorical claims" of both approaches are central to the way in which they present their implications for policymaking. The "rhetorical claims" in behavioural economics are noted when their proponents emphasise that nudges promote subjective wellbeing and correct rationality failures. The "rhetorical claims" in fast and frugal research are noted when their proponents emphasise that boosts are better because they empower decision-makers. Normativism affects how the two approaches inform policymaking. The policy goal of nudges is to correct decisions; the policy goal of boosts is to improve agents' capacities to make good choices.

Highlighting the normative dimension of the debate between the heuristics and biases research programme and the fast and frugal heuristics programme and especially between their policy approaches suggests that the problems with the way in which behavioural economics is used in policymaking are related more to its normative recommendations than to the implementation of nudges. As a result, the problems that the fast and frugal heuristics approach identifies with the main findings from behavioural economics are relevant when it comes to looking at how behavioural economics can inform policymaking but do not invalidate nudges' value as a policy tool. Overall, the main lesson is that normative claims should be avoided, but nudges are still practical interventions for pursuing behavioural change.

Firstly, the fast and frugal approach emphasises that behavioural economics cannot support the fact that nudges generally promote subjective wellbeing (Gigerenzer, 2015). I agree with this assessment. In chapter 2, I have already discussed the idea that the claim that nudges improve subjective wellbeing cannot be generally supported, and that such a claim is problematic. As a result, it should be avoided. However, I have also pointed out that nudges still have value as policy tools when they are used to promote other policy goals.

Secondly, the fast and frugal approach emphasises that behavioural economics cannot support the claim that people lack rationality and that agents cannot learn to overcome their biases. Gigerenzer (2018) claims that these arguments are problematic because, based on recommendations from behavioural economics, policymakers that want to apply strategies to promote agents' subjective wellbeing may select nudges instead of boosts or other educational strategies. I also agree with this assessment. It is unclear why people sometimes make choices that undermine their own wellbeing. Cognitive biases might be a reason, but lack of information, lack of opportunities and lack of competences might also be relevant; therefore, educative interventions should also be considered. However, it should also be noted that boost policies should not be preferred on the basis that they are more deliberative and educational. The decision on whether to use a nudge or a boost to improve decision-making should be based on which interventions are expected to perform better<sup>15</sup>.

Finally, regarding the use of nudges, the criticism from the fast and frugal heuristics approach does not claim that the factors identified in behavioural economics are not relevant enough to be used to change behaviour. The effects of mechanisms identified in behavioural economics are sufficiently prevalent to be relevant for their use in seeking behavioural change. The criticism from the fast and frugal heuristics research programme is rather that their implications are not as negative and irreparable as conceived by behavioural economics.

## **5. Concluding remarks**

The theoretical debate on rationality and decision-making leads to important considerations about how behavioural insights are used in policymaking. The ways in which behavioural insights inform policymaking are influenced by the positive findings of the research programmes but also by their normative assumptions and claims on rationality and decision-making. As a result, it is necessary to take the research programmes underlying the main policy recommendations into account in order to understand the way in which behavioural insights can contribute to policymaking and the challenges and limitations they face.

---

<sup>15</sup> There are already some interesting studies comparing their performance. Bradt (2019) compares the effectivity of boosts and design of nudges to increase consumer demand for insurances, and Franklin et al. (2019) compare their effectivity at improving financial decisions.

The application of behavioural insights in the last decade has progressed with little reflection on how they are used. Behavioural economics develops policy recommendations that overstate the idea that people are irrational and that biases are difficult to overcome. In practice, this is problematic because these recommendations are not fully supported by empirical claims but are presented as such, leading to regulatory problems and problems of effectiveness. The objection from a fast and frugal heuristics approach highlights these negative implications and correctly points out that insights in behavioural economics should not be used to justify a policy agenda. However, the fast and frugal heuristics approach faces similar problems by rejecting utility theory and its commitment to the study of “ecological rationality”. Firstly, criticism of this approach tends to overstate the negative implications of behavioural economics in policymaking and present all their recommendations as problematic. Secondly, its own approach to behavioural change also makes somewhat unfounded normative claims.

Promoting policies based on specific conceptions of rationality and decision-making, regardless of whether they are interpreted as positive or negative, is problematic. Policy interventions should not be justified according to specific approaches to rationality. In general, it cannot be said that all judgments have proven to be systematically bounded and always incorrect; nor does the judgment and behaviour that comes from ecological heuristics always yield a good answer. Therefore, none of the approaches should be used to justify a normative agenda in policymaking. The goal should be to explore how the recommendations from both research programmes contribute to designing better policy interventions when given some normative foundations, rather than addressing which programme provides a better theoretical foundation for behavioural public policy.

Insights on human rationality and decision-making may help to devise better public policy. However, lessons should be practical and not normative and should be intended to inform about behavioural tendencies, rather than provide a complete interpretation of human decision-making. A lesson can be drawn from how rational choice theory is relevant when it comes to informing policymaking. Rational choice provides very valuable insights for policymaking. Self-interest is a relevant motivation for action, which plays an essential role in many problems related to collective action. Likewise, incentives, sanctions, and information are crucial assets in the pursuit of behavioural change and tend to be effective in many domains. However, problems arise when claims about rational choice are used to justify the premise that self-interest is the only relevant motivator, and when incentives are preferred and applied indistinctly with no reflection on their possible negative implications. The same lessons apply to behavioural insights. The recommendations from behavioural insights are valuable, recognise the relevant motivations for actions and provide interesting ideas on how to nudge or boost people to change their behaviour. However, their claims should not be used without taking their limitations into account.

## Chapter 4

### A pragmatic definition of nudges

#### 1. Introduction

As discussed in chapter 1, the original characterisation of nudges defines them as a unitary concept with five properties. A nudge:

- i. does not change economic incentives
- ii. does not forbid any options
- iii. affects the behaviour of Humans and is ignored by Econs
- iv. improves people's wellbeing as judged by themselves and
- v. respects freedom of choice.

With these properties in mind, Thaler & Sunstein (2008) defined nudges as a policy instrument that has practical and normative value to achieve behavioural change. However, as discussed in Chapter 2 and Chapter 3, their understanding of nudges entails several problems and leads to doubts about the acceptability of nudges as a policy tool. In the last two chapters, I have addressed some of the most important objections in relation to using nudges in policymaking. Responding to these objections, I have defended the pragmatic value of nudges as policy tools. However, as noted in Chapter 1, my arguments are in favour of nudges, but not in favour of the original definition of them. Thaler and Sunstein's idea of nudges has significant problems and, therefore, I suggest that the definition of nudges should be revised.

Defining nudges with the five original properties is problematic in two main ways. Firstly, the original definitional properties assign nudges normative content based on limited empirical evidence. The claims that nudges improve people's subjective wellbeing and respect agents' freedom of choice are based on normative assumptions and lack empirical evidence. Secondly, the original characterisation of nudges is also problematic because it presents them as a unitary concept. To work as intended, nudges ought to comply with properties (i), (ii), (iii), (iv) and (v). While the properties are not contradictory, they rarely hold together. As a result, the concept of nudges has an ambiguous meaning, as it is not clear which of the stated properties is more important as far as defining whether an intervention is a nudge or not. It is not at all clear and hinders the discussion on the ways in which nudges can contribute to policymaking.

I argue that a more pragmatic definition of nudges is needed to advance the debate on whether nudges are acceptable policy tools, discuss their ethical implications and explore their performance in policymaking. Accordingly, I present a new definition of nudges that proposes dropping normative properties (iv) and (v) and redefines property (iii). The definition is intended to strip nudges from their

normative claims and define them according to the features that are essential when it comes to understanding the ways in which nudges influence people's choices.

Chapter 4 is a brief conclusion of the arguments discussed in chapters, 1, 2 and 3. The chapter discusses which properties in the original characterisation should be either dropped or redefined and presents the new definition. The chapter ends with some concluding remarks that emphasise the importance of complementing the theoretical discussion on the acceptability of nudges with empirical work on their public support.

## **2. Problems with some of the original properties**

The first fundamental problem with the original characterisation of nudges is the normative content. The original characterisation of nudges assigns normative qualities to nudges. Nudges are not only conceived as tools to change behaviour but as tools that promote positive values and ends. In this sense, nudges are unlike any other policy instrument, because other measures aimed at changing behaviour do not have associated normative goals. However, properties (iv) and (v) are the source of several objections to nudges and have monopolised the debate about the implications of nudges as policy tools. As a result, and in line with the arguments in chapters 2 and 3, I suggest that nudges are more useful and less ethically problematic when their use is not associated with any normative ideas.

Thaler & Sunstein (2008) define nudges as interventions that respect agents' freedom of choice and emphasise that this is an essential feature when it comes to supporting their implementation. However, as argued in Chapter 2, property (v) is very problematic. Overall, it appears clear that nudges cannot generally be justified on the basis that they respect agents' freedom of choice. Even though nudges do not modify economic incentives and maintain all the available options in a context of choice, they interfere in people's decision-making and may diminish agents' ability to make their own choices. As a result, properties (i) and (ii) are insufficient to ensure that nudges protect agents' freedom of choice. However, I do not see any problem with the fact that nudges might not promote freedom of choice. As also suggested in Chapter 2, protecting freedom of choice might be relevant in some situations but may be irrelevant or even have harmful consequences in others. As a result, the ethical value of nudges should not be based on the idea that they must respect freedom of choice.

The other normative quality of nudges is expressed in property (iv) and emphasises that nudges are interventions that promote agents' subjective wellbeing. Again, as argued in chapters 2 and 3, objectors point out problems with this claim. Overall, the consensus is that the claim is supported on limited empirical evidence and, therefore, it cannot be assumed that nudges always promote subjective wellbeing. As in the case of property (v), property (iv) leads to criticism against nudges and limits the possibilities of using nudges to promote goals beyond individuals' subjective wellbeing. Accordingly, I suggest that nudges should not be defined as interventions that promote a specific policy goal.

Research on behaviour and decision-making suggests that policymakers should interfere in people's choices to promote their wellbeing in some cases. However, there is no reason why nudges should only be used to this end. As argued in Chapter 3, evidence from research on decision-making beyond nudges also supports the use of boost policies designed to improve people's capacities to make choices. Likewise, traditional interventions, such as incentives or information and educational campaigns, may also be useful for these purposes. Nudges might be used to promote subjective wellbeing, but they cannot be generally understood as interventions that do so. Again, I see no problem with not endorsing property (iv). As argued in Chapter 2, nudges have the potential to promote different outcomes and can be very beneficial when used to tackle social problems. Given the problems of nudges in being truly pro-self and their potential to promote pro-social goals, it would be more beneficial not to define nudges as interventions that promote a specific goal but to explore their potential to resolve social issues instead.

Rather than defining nudges as intrinsically ethically permissible interventions because they promote subjective wellbeing and respect freedom of choice, the normative implications of nudges should be addressed by considering whether nudges are being used to promote relevant and legitimated outcomes. Overall, dropping properties (iv) and (v) is useful in three main ways. Firstly, it is useful for overcoming some of the objections to nudges. Secondly, it allows the discussion to move beyond the original normative justification of nudges and explore the cases in which their implementation is permissible without preconceived ideas about which ends and values they should promote. Finally, it presents nudges as valuable policy tools to promote policy outcomes beyond the ones defined in their original nudge characterisation.

The last problem with the original definition of nudges is property (iii). Thaler & Sunstein (2008) define nudges as interventions that affect the behaviour of Humans and are ignored by Econs. They use this definitional property to identify the theoretical insights of the approach and to describe nudges as interventions that relate to psychological mechanisms in the categories of biases and blunders, temptation and following the herd. However, the wording of property (iii) is problematic because it expresses positive elements but also normative assumptions.

Defining nudges using the distinction between Humans and Econs could create confusion and might raise objections that which could be avoided with a more explicit reference to the mechanisms that nudges are expected to tap into. Firstly, the distinction between Humans and Econs does not identify which mechanisms underpin the use of nudges. This contributes to the ambiguity of the concept and its misuse. Secondly, the wording favours the interpretation of the psychological mechanisms used by nudges as behavioural biases, which, as argued in Chapter 3, is problematic. Finally, property (iii) might suggest that the implementation of nudges is only informed by behavioural economics. However, as argued in chapters 2 and 3, evidence can be found across social sciences on the effects of the factors under the biases and blunders, temptation and following the herd categories. For instance, research on

social influence might not be committed to the notion of rationality and decision-making that underlies research on behavioural economics. However, it might present relevant insights on how to exploit the ‘following the herd’ tendencies for policymaking. Therefore, I suggest explicitly defining nudges as interventions that appeal to mechanisms related to biases and blunders, temptation and following the herd.

### **3. A pragmatic definition of nudges**

As per the discussion in chapters 1, 2 and 3 and the considerations outlined in the previous section, I suggest that the term ‘nudge’ should be used to reference interventions that comply with the three definitional properties. A nudge:

- i. does not change economic incentives
- ii. does not rule out any options
- iii. affects the behaviour of agents by appealing to factors related to biases and blunders, temptation and/or following the herd.

However, the term ‘nudge’ should be avoided when referring to interventions that do not comply with the three main definitional properties, even though they may (a) improve people’s wellbeing as judged by themselves and (b) respect agents’ freedom of choice.

The definition trims down the normative implications and assumptions of nudges and is based on what is essential when it comes to understanding the ways in which nudges influence people’s choices and what makes them different from other interventions seeking behavioural change.

### **4. Final remarks**

In the first part of the thesis, I discussed the debate around the acceptability of nudges as policy tools and focused on ethical, theoretical and conceptual issues. I highlighted two main themes to address the main objections and concerns regarding nudges. I emphasised that critics are right in pointing out the problems with the original characterisation of nudges. However, I also suggested that nudges are valuable policy instruments that are not problematic per se when we look beyond the original justification for using them. However, several questions on the acceptability of nudges remain to be answered. Likewise, while the theoretical discussion provides relevant ideas about the ways in which nudges should be applied, the philosophical debate is inconclusive.

The discussion on the role and acceptability of nudges as policy tools should not only be informed by the theoretical discussion but also by people’s opinion of nudges. Exploring the public acceptance of nudge interventions is important when it comes to complementing the ethical and theoretical discussions and ensuring that nudges have their intended effect. Firstly, for nudges to be successful, people need to approve of their use. If people perceive nudges as an ill-intended attempt to manipulate

their behaviour, the interventions might backfire. Consequently, the perception and acceptance of these measures by citizens is an important factor for ensuring their effectiveness. Secondly, understanding how people perceive nudges may inform the philosophical discussion on the concept. The ethical debate is dominated by abstract notions and philosophical ideas that express conflicted views; as a result, neither proponents nor opponents of nudging can claim to clearly have the upper hand. While knowing how people view nudges does not provide a conclusive answer to the debate, complementing the theoretical discussion with people's views may advance the conversation towards a more pragmatic focus: i.e., the implementation of nudges in policymaking.

The following chapters move away from the theoretical discussion to an empirical exploration of how people feel about the implementation of nudges. Chapter 5 introduces the tendencies and main findings from studies on public support for nudges. Chapter 6 and 7 provide evidence about how people feel about the use of different nudges in different policy domains. These three studies are intended to contribute to the current discussion on attitudes towards nudges and provide answers to the central theme of this thesis: i.e., the debate on the positive and negative implications of using nudges in policymaking.

# Chapter 5

## Attitudes towards nudges: a literature review

### 1. Introduction

The application of nudges has proven to be effective in a wide range of policy domains and when used to promote different policy goals around the world. Despite its growing popularity, its use remains a controversial topic. As discussed in Chapter 2, the strategy faces objections regarding its ethical legitimisation. As discussed in Chapter 3, it also faces complaints regarding its limitations as a behavioural change intervention. Both debates have been prominent in academic literature, and, while the literature has established relevant arguments for and against nudges, some questions remain unanswered.

An interesting insight into tackling both issues is to consider how citizens feel about nudges. Knowing how people feel about nudges is relevant to informing and shaping the government's policy choices, ensuring that citizens will accept them, and improving their effectiveness and legitimisation. Recently, research on this topic has become more prominent. Current studies take into consideration questions about nudges' effectiveness and ethical considerations. Some studies focus on the big picture and study the general acceptance of nudge interventions, considering preferences for different types of nudges, preferences for nudges with different goals, and exploring which individual, socioeconomic and geographic factors affect these attitudes. These studies aim to comment on the political feasibility of nudges by exploring whether citizens approve of their use. Other studies focus on ethical questions, especially trying to understand under which circumstances people accept the intrusiveness of nudges on their choices and how they perceive their effect in terms of autonomy, wellbeing and freedom of choice.

Research indicates that people generally approve of the use of nudges, but different factors affect whether they will support or oppose their implementation. The evidence available has both positive and negative implications for using nudges in public policy. The general acceptance of nudges paints a positive picture for implementing them, contributes to their legitimisation, and supports the notion that nudges are feasible tools for changing behaviour. Likewise, the positive findings act as a counterweight to the philosophical discussion in which nudges tend to be viewed negatively. By contrast, the confirmation that some interventions spark more disapproval than others, and the fact that many different factors seem to affect preferences towards nudges trigger concerns about their use. These results suggest that the idea of *one-nudge-fits-all* is unattainable. Overall, we need more evidence about the factors that affect nudge acceptance and what these findings imply for the implementation of nudges in policymaking.

Chapter 5 presents a literature review on studies on the public acceptability of nudges. Section 2 details the characteristics of the studies on attitudes towards nudges and the main findings. Section 2.1 summarises the methods and features of the main studies; section 2.2 summarises the main results. Section 3 discusses gaps and issues that are still unresolved and the questions that will be explored in the following empirical chapters. The chapter ends with a few conclusions.

## **2. Attitudes towards nudges: an overview.**

In recent years, studies on public attitudes towards nudges have become more prominent. Several studies have explored whether people like nudges, which nudges they favour and in which areas their use is accepted. Studies have also examined why people approve or disapprove of nudges and which factors affect people's acceptability of nudge interventions.

Research on attitudes towards nudges is still relatively new and limited, but studies have reported a few consistent findings. Firstly, people generally approve of the use of nudges as policy tools. Approval ratings vary significantly, but nudges are usually supported by the majority. Secondly, people reject nudges when they disagree with their ends. Overwhelmingly, disapproval of nudges comes from disliking the nudge policy outcome or the party or politician that implements the intervention (Loibl et al., 2018; Sunstein, 2017b; Tannenbaum et al., 2017). Both results indicate that, in general, people have positive attitudes towards the implementation of nudges and value them as they would do with any other policy instrument.

At the same time, research also reports issues that are specific to nudges. Firstly, people reject nudges when they perceive that they have illicit ends, for instance, when nudges contradict individual and social values or social norms (Sunstein, 2017b; Tannenbaum et al., 2017). Secondly, people generally reject nudges that go against individual economic interest (Hagman et al., 2015; Jung & Mellers, 2016; Sunstein, 2017b). Thirdly, people intuitively differentiate nudges according to their degree of intrusiveness and tend to favour nudges that are more educational, transparent and engage deliberation over nudges that limit reflection and are less noticeable (Felsen et al., 2013; Jung & Mellers, 2016; Sunstein, 2017c). Finally, behavioural, psychological and sociodemographic factors affect the approval nudges of in specific ways (Hagman et al., 2015; Jung & Mellers, 2016; Sunstein et al., 2019).

Findings have shown consistent tendencies, but many questions remain unanswered. Given the increasing interest of governments in using nudges, more research on their public acceptability is needed to clarify how people form opinions about nudges and what implications these findings have on their use in policymaking. In the sections below, I have reviewed studies on attitudes towards nudges. In an attempt to provide a summary of the literature that is useful in guiding empirical research, the following sections report the characteristics of the studies on attitudes towards nudges and the main findings. Section 2.1 summarises the methods and features of the main studies; section 2.2 summarises the main results.

## **2.1. Overview of the methods and characteristics**

Empirical research on attitudes towards nudges has used surveys as well as experimental methods, in the form of surveys-experiments and vignettes-experiments. Surveys are a good method for gathering large samples and obtaining a general impression of how people feel about nudges. Likewise, surveys show how different factors influence people's views on the matter and are useful for making cross-country comparisons. Experiments are better at identifying causal factors. The experimental context makes it possible to manipulate independent variables (treatments) to determine how they impact on dependent variables. Experiments make it possible to identify causal links between the different factors and people's attitudes towards nudges. The results of surveys and experiments are both relevant and complementary. As a complement to the literature on attitudes towards nudges, some studies analyse the effect of disclosing nudges. In some studies, where lab and field experiments are used, authors have investigated how people feel about their decisions when they are informed about the presence or the effect of a nudge (Bruns et al., 2018; Kroese et al., 2016; Loewenstein et al., 2015). These studies have also indirectly indicated how people feel about nudges and about being nudged.

The majority of surveys on attitudes towards nudges have been conducted in the United States and Europe. Jung & Mellers (2016) and Sunstein (2017b) have conducted surveys in the United States. Evers et al. (2018) reported findings from the United States, Germany, The Netherlands, France, Italy, Poland, Bulgaria, and the United Kingdom. Similarly, in a series of replicated surveys, Reisch & Sunstein (2016) looked at Denmark, France, Germany, Hungary, Italy, and the United Kingdom, while Loibl et al. (2018) explored attitudes towards nudges in Denmark, Hungary, the United Kingdom and Italy, and Hagmann et al. (2018) studied attitudes towards nudges in Switzerland.

Some studies have extended the scope and explored comparisons between countries outside Europe and the United States. In 2018, Sunstein et al. explored attitudes towards nudges in Australia, Brazil, Canada, Japan, China, Russia, South Africa, and South Korea. In 2019, Sunstein et al. explored attitudes towards nudges, comparing the results of their survey in Germany, Denmark, South Korea, the United States and Flanders (Belgium). The majority of surveys have been conducted online and employed large and nationally representative samples by gender and age. Likewise, some of the studies have used the same questions, which facilitates cross-country comparisons.

While employing smaller samples, some experiments have also used samples from different countries. Arad & Rubinstein, (2018) compared results from samples in Germany, the United States and Israel. Felsen et al. (2013) compared results from respondents in the United States and Canada. Hagman et al. (2015) conducted an online experiment with contrastive vignettes using samples from Sweden and the US, and Petrescu et al. (2016) used large samples from the United Kingdom and the United States. Current studies on attitudes towards nudges are robust, include different countries and allow interesting

comparisons. However, more research needs to be conducted, and it would be beneficial to expand the research into other countries.

In all the studies on attitudes towards nudges, the primary dependent variable is people's approval of nudge interventions. To capture support for nudges over different interventions, some studies have used a simple bivariate response in which people have to choose between approving or disapproving of nudges (e.g., Loibl et al., 2018; Sunstein, 2017b; Sunstein et al., 2019). Other studies have employed Likert scales with more than two points that allow subjects to express how much they agree or disagree with a particular nudge (e.g., Felsen et al., 2013; Jung & Mellers, 2016).

Beyond support for interventions, a majority of studies have also collected people's perceptions of other aspects of nudges. Notably, some studies have collected people's views on the effectiveness of nudges (e.g., Evers et al., 2018; Petrescu et al., 2016). It has also been popular to consider whether nudges represent a threat to people's freedom of choice or autonomy, whether nudges are perceived as intrusive, whether nudges involve manipulation, whether interventions are paternalistic, and whether people see nudges as positive tools for promoting individual and/or societal wellbeing (Felsen et al., 2013; Hagman et al., 2015; Jung & Mellers, 2016; Niker et al., 2018). All of these perceptions have then been analysed as co-variables, to further clarify which factors affect people's acceptability of nudges.

Research on attitudes towards nudges has included several co-variables. There has been significant interest in exploring how various factors might affect people's opinions of nudges. Several studies have collected information on sociodemographic characteristics, such as gender, age, education, income and region. Also, a large portion of studies have asked questions about political affiliation and political views, and people's trust in institutions or people's opinions on who applies the nudges (Loibl et al., 2018; Sunstein, 2017b; Sunstein et al., 2018, 2019). As other relevant co-variables, some studies have included questions on people's behaviour and lifestyle choices, especially in terms of the actions that nudges target; for instance, whether people are organ donors or whether and how they consume specific products, such as alcohol and tobacco. Some studies have also included questions on people's personal worries regarding the issues that nudges intend to change, for instance, worries about personal health and concerns about the environment.

Interestingly, some studies have also explored individual dispositions. For instance, Jung & Mellers (2016) measured individuals' empathy, individualistic views, desire for control and psychological reactance. Niker et al. (2018) collected information on the character of individuals in terms of their communitarian and individualistic views. Finally, some studies have explored people's subjective perceptions of self-control, in general, and in specific domains, as well as opinions of how much others enjoy self-command in different areas (Gyrd-Hansen & Kjær, 2015; Hagman et al., 2015; Petrescu et al., 2016).

With the exception of studies that have explored the consequences of disclosing nudges, studies on attitudes towards nudges have tended to present nudges as hypothetical or potential future policies. If some of the nudges have been implemented, studies have not referred to any aspect of real interventions. Using hypothetical scenarios facilitates research and is useful for collecting opinions on different topics; however, what respondents think about a hypothetical situation may differ from what they would feel in the real case scenario. Consequently, it could be interesting to incorporate questions about real policies and develop more experimental research on how people feel about nudges when they are informed about their presence and the impact they had on their decision-making.

All of the studies have avoided using the term nudge and instead opted for describing policy instruments in an understandable and simple manner. Junghans et al. (2015) found that people were generally unfamiliar with the term nudge and understood it better when they were presented with examples. They also found that people who knew the term were sometimes also familiar with the debate around their implementation, which could have compromised their responses. The consensus is that it is better to avoid the explicit use of the term nudge.

A particularly interesting issue to look at when looking at attitudes towards nudges is the type of nudges and the areas and goals that nudges have. As previously discussed, the term nudge references many different interventions with different degrees of intrusiveness. Likewise, nudges can be applied to promote different policy goals, both pro-self and pro-social, and can work in many different areas. As a result, when addressing attitudes towards nudges, it is essential to consider which nudges-interventions and types of nudges frequently appear in studies and in which areas their use is being evaluated. Both factors shape preferences towards nudges, and it is crucial to collect opinions on different nudges and their use in various domains. Otherwise, studies could report biased results that only reflect people's attitudes towards particular nudges and their use in specific areas.

Table 6 provides a summary of the areas and types of nudges that different studies include. The table lists the domains in each study, distinguishing between pro-self and pro-social, and describes the interventions included in each study with generic labels.

**Table 6. Domains and types of interventions considered in the attitudes towards nudges literature.**

<b>Reference</b>	<b>Domains</b>	<b>Interventions</b>
Arad & Rubinstein (2018)	Pro-self: retirement savings healthy eating	Default layout of options automatic enrolment
Davidai & Shafir (2018)	Pro-self: health coverage bill payments on time retirement savings	default settings information
Evers et al. (2018)	Pro-self: healthy eating	placement sizing
Felsen et al. (2013)	Pro-self: healthy eating purchasing behaviour exercising savings productivity	placement labels layout of options feedback (anchoring a total budget giving information of time spent on the internet) information
Gyrd-Hansen & Kjær (2015)	Pro-self: child immunisation alcohol consumption dental check-ups healthy eating general medical visits smoking participation in screening programmes exercising Pro-social: organ donation	bans tax leveraging attractiveness of small economic incentives
Hagman et al. (2015)	Pro-self: smoking healthy eating Pro-social: organ donation climate compensation energy consumption tax evasion	default settings social comparison (descriptive social norms communication) communication of prescriptive social norms re-design of physical environment labels graphic image warnings
Hagmann et al. (2018)	Pro-self: healthy eating	placement tax bans sizing labels substitution of sugar by artificial sweeteners educational campaigns

<b>Reference</b>	<b>Domains</b>	<b>Interventions</b>
Jung & Mellers (2016)	Pro-self: health coverage privacy retirement savings healthy eating and health-related behaviour smoking personal expenses voting Pro-social: organ donation driving behaviour charity donations water conservation	automatic enrolment default settings layout of options placement optical illusions option attractiveness graphic image warnings information with feedback alerts notifications/reminders
Junghans et al. (2015)	Pro-self: healthy eating	location placement
Niker et al. (2018)	Pro-self: healthy eating voting	social influence subliminal images information
Petrescu et al. (2016)	Pro-self: healthy eating	educational campaigns location sizing shape tax
Reisch & Sunstein (2016) Loibl et al. (2018) Sunstein et al. (2018) Sunstein et al. (2019)	Pro-self: healthy eating smoking driving behaviour Pro-social: environmental protection organ donation charity donations	labels (informative and traffic-light system) automatic enrolment prompt choice placement educational campaigns subliminal advertisements default settings information bans location
Sunstein (2017b)	Pro-self: healthy eating driving behaviour smoking savings voting Pro-social: environmental protection organ donation charity donations anti-discrimination	labels (informative and pictures) information graphic warnings automatic enrolment financial literacy program educational campaigns default settings aggressive campaigns consultancy sessions Prompting choices placement educational campaigns subliminal advertisements bans anchoring location

Reference	Domains	Interventions
Sunstein (2017c)	Pro-self: smoking retirement savings healthy eating voting abortion Pro-social: environmental protection water conservation	information graphic warnings automatic enrolment financial literacy program educational campaigns default settings aggressive campaigns consulting sessions
Tannenbaum et al. (2017)	Pro-self: retirement savings access to assistance programs access to tax breaks safe sex intelligent design	automatic enrolment planning prompts leveraging loss aversion prompt public commitments descriptive social norms
Yan & Yates (2019)	Pro-self: retirement savings Pro-social: organ donation environmental protection	opt-out policies

In terms of areas, the majority of studies have explored attitudes towards pro-self nudges. As seen in Table 6, there seems to be an overrepresentation of specific areas. Several studies that have explored attitudes towards pro-self nudges have focused on health-related behaviour and financial decisions. Sixteen out of the nineteen studies included questions on interventions in the domain of healthy eating, especially regarding the consumption of sugar; nine included interventions to reduce smoking, and two included questions on interventions to promote physical exercise. Likewise, six out of nineteen studies included nudges aimed at improving retirement savings, and four studies included some sort of decisions related to savings or budgeting for personal expenses.

While less frequent, some studies have also investigated how people feel about nudges that promote pro-social aims. Again, with regard to pro-social domains, there are specific areas that have appeared frequently. Organ donation has been the most common domain; nine out of the nineteen studies have asked about the use of nudges to promote organ donation. Next, the most repeated area has been environmental protection; seven of the studies have included questions about the use of nudges in that domain, and three studies have included questions on related issues, such as water conservation, energy consumption and carbon offsetting. Finally, another recurrent area has been charity donations; five studies have asked about that topic.

The most frequent pro-self and pro-social domains are all mentioned in *'Nudge'* and are areas in which behavioural factors tend to affect decisions and in which the use of nudges has proven to be useful and effective. The focus on pro-self domains and interest in the use of nudges in such areas also goes back to the original nudge characterisation, and the idea that nudges should promote individuals' wellbeing

and make people healthy and wealthy. However, the review indicated that views on the use of nudges in other relevant areas still need to be studied more. For instance, literature is increasingly exploring the use of nudges to promote tax compliance, yet only one of the studies has asked about that specific domain. Likewise, as mentioned in the previous chapters, given the problems of nudges to be truly pro-self and their potential to promote pro-social goals, it would be beneficial to focus more on pro-social domains and broaden the areas taken into consideration.

Studies on interventions also have similarities. In general, studies have tended to combine questions about nudges and traditional interventions, especially taxes, mandates, the provision of information and the use of educational or informative campaigns. Regarding no-nudge interventions, the majority of studies have tended to include questions on mandates, taxes, and many have included questions on subliminal messages. The addition of these interventions allows us to see how people distinguish and compare nudges to traditional interventions and more intrusive interventions.

The most frequent nudge in studies about public support for nudges has been the use of default settings, such as automatic enrolment and opt-out policies. Eight out of nineteen studies have included questions about automatic enrolment in different programmes and registers, and nine have included questions on default options. The second most frequent nudge has been the placement or change of location of products. Ten studies have included questions about the change of placement of different products to either facilitate or impede access to specific options, and four out of nineteen have included questions about changing the location of products.

Other recurrent interventions have been the use of labels and warnings, interventions that alter the size of products and interventions that alter the layout or attractiveness of options. Eight studies have included questions about labels, four have included questions about the layout and attractiveness of options, and four have included questions on altering the size of specific products. Finally, while less frequent, some studies have included questions about prompting choices or plans, leveraging loss aversion, anchoring information, using feedback, using social influence and social norms and using optical illusions.

Several studies have included questions about information and education campaigns. In many studies, these interventions have been labelled as nudges; however, there is significant controversy over whether they should be considered nudges. By contrast, interventions that are a better fit with the definition of nudges are underrepresented. For instance, only four studies have included questions about nudging using social norms (descriptive and prescriptive) or social influence. Similarly, questions about using framing effects are also quite uncommon. The absence of questions regarding these interventions might be due to the fact that explaining them to the subjects is more complicated than it is with other interventions. However, as with domains, future studies should expand the type of nudges included in the questions.

A final aspect that is of interest in studies on attitudes towards nudges is how they present the nudger. In the majority of studies, governments or different figures in public administration have been announced as the ones implementing the nudge. In these cases, the reference has been neutral. However, some studies have explored how people's attitudes towards nudges change when they come from other sources. Evers et al. (2018) compared views on nudges when applied by policymakers, experts and companies. Likewise, others have studied support for nudges when the employer applies the nudges (e.g., Davidai & Shafir, 2018; Felsen et al., 2013; Jung & Mellers, 2016). While it is crucial to studying how people will respond to nudges when applied by public administrations, exploring how people respond to nudges when applied or endorsed by other sources is also relevant to discussing alternative ways to implement and communicate nudges.

To conclude this section, the overview of the methods and characteristics in literature on attitudes towards nudges has shown consistencies between studies. The use of surveys and experiments are adequate methods and provide complementary results, and the development of studies in different countries is also positive. Studies have also been consistent in how they have described nudges and have tended to include: a) references to similar scenarios, b) questions about the same nudges and areas of application and c) interesting and diverse co-variables. All of these elements have provided a solid base for future studies.

However, some aspects need to be revised to advance the research. Firstly, studies should contain more questions about nudges that are unrepresented, particularly social-norm nudges, the use of framing and the use of self-control devices. Secondly, studies should include more questions on the use of nudges in pro-social domains and include areas in which the potential to include nudges is high but tends to have not been studied much, for instance, tax compliance. We know quite a lot about how people think about nudges applied to healthy eating and how people feel about default settings. However, we lack information on how people react to other nudges and how they value their use in other domains. Finally, to go beyond the limitations of asking about hypothetical scenarios, it could be interesting to include more questions about real interventions and to dig deeper into how people feel about nudges when nudges impact on their real decisions.

## **2.2. Overview of the findings**

The most common and positive finding across studies on attitudes towards nudges is that the acceptance of nudges is generally high. Several studies indicate that people approve of nudges, in different forms and applied to different domains. While the acceptance of nudges varies significantly depending on the intervention and the aim, only very few interventions are generally opposed.

- *General findings*

In a series of surveys replicated in certain countries, Sunstein (2017b), Reisch & Sunstein (2016), Loibl et al. (2018) and Sunstein et al. (2018) found high acceptability for several nudges. Support for the majority of interventions oscillated between 50% and 90%, and many had acceptability rates around the 70-80% mark. With the exceptions of Japan, Denmark and Hungary, i.e., countries with lower overall support for nudge interventions, the acceptability of nudges was high. Likewise, nudges tended to have more support than conventional and more restrictive policy tools like taxes, bans and mandates. In general, the results tended to show overall public support for the implementation of nudges.

Turning now to specific nudges, literature has reported consistent results on the acceptability of many interventions. For instance, the use of label-nudges has been broadly supported, particularly when applied to promoting healthy eating, but also to providing information about other aspects (Evers et al., 2018; Hagmann et al., 2018; Jung & Mellers, 2016; Loibl et al., 2018; Sunstein, 2017b). Similar nudges, such as using posters for displaying relevant information within the context of choice and using graphic warnings on cigarette packets have also received great support (Hagman et al., 2015; Jung & Mellers, 2016; Sunstein, 2017b). These positive results indicate that in general, labels and other forms of displaying information close to the context of choice are likely to be accepted by the majority.

Information and educational campaigns to promote both pro-self and pro-social goals are also supported. For instance, Sunstein (2017b) found great support for educational campaigns to promote healthy eating, reduce smoking and reduce distracted driving and found that information and educational campaigns were only opposed when they promoted problematic or controversial aims.

Similarly, interventions that provide feedback have been well received. In a nationally representative survey in the US, Jung & Mellers (2016) found that a majority of respondents approved of using spending alerts, providing clear information about people's projected monthly income in retirement according to their present levels of savings and using notifications to voters with information about how to get to polling stations. Along similar lines, Tannenbaum et al. (2017) found support for interventions that promoted choices and facilitated planning. In a condition of the survey experiment in which they asked about different nudges without explaining their aim, the authors found support for using plan-making prompts (average of 3.53 on a Likert scale from 1 to 7) and prompted public commitments (average of 3.16 on a Likert scale from 1 to 7).

Some studies found support for interventions that modified the attractiveness of options either by changing their disposition or anchoring specific information. For instance, Felsen et al. (2013) found support for interventions that use these techniques to improve savings, by placing less needed items at the bottom on an online marketplace, and by displaying the available budget on the sidebar. Similarly, Tannenbaum et al. (2017) found support for leveraging loss aversion as an intervention to change behaviour (average support of 3.27 points on a Likert scale from 1 to 7).

Interventions that change the placement, location and size of products to promote healthy choices are also broadly supported. In Sunstein (2017b), Reisch & Sunstein (2016), Loibl et al. (2018) and Sunstein et al. (2018) the placement of healthy foods in a prominent and visible location and the implementation and use of ‘sweet-free’ cashier zones in supermarkets received high levels of support (average approval rating of 72% and 62%, respectively). Jung & Mellers (2016) found that a majority supported the placement of healthy foods in suitable locations in cafeterias. In a nationally representative survey in Switzerland, Hagmann et al. (2018) also found considerable support for reducing the availability of unhealthy products by changing their location (average of 4.83 on a scale from 1 to 7). In a survey experiment conducted in the US and the UK, Petrescu et al. (2016) found an average approval rating of 67% for interventions that changed the location of unhealthy sugary drinks; an average approval rating of 59% for only selling sugary drinks in small containers, and average support of 52% for an intervention that changed the shape of sugary drink containers to nudge people into drinking less. Likewise, Evers et al. (2018) found support for making unhealthy snacks less visible and less available (average support of 5.38 and 5.41 on a scale from 1 to 7) and support in several countries for reducing portion sizes to reduce consumption (average support of 4.61 on a scale from 1 to 7). Also, Hagman et al. (2015) found average approval ratings of above 80% in both Sweden and the US for re-designing cafeterias to promote healthy choices. Healthy eating is one of the most studied domains, and many studies have examined attitudes towards these same nudges. The results confirm that people support these types of interventions.

The use of social nudges, i.e., nudges that communicate descriptive or prescriptive norms and employ social influence has been generally supported. Loibl et al. (2018), Reisch & Sunstein (2016) and Sunstein, (2017b) found high levels of support for using social comparison to promote energy conservation (average approval rating of 66.4% and 67.1% in Sweden and the US respectively) and even higher levels of support for the use of prescriptive moral messages to promote tax compliance (average approval rating of 81.5% and 80%). However, Tannenbaum et al. (2017) asked respondents how they felt about using social comparison without explicitly stating to what end it would be used. In this case, a majority of respondents opposed the intervention (average of 2.95 on a Likert scale from 1 to 7). Studies on the use of social norm nudges are still uncommon, so these results are very much explorative.

Public support for any kind of default settings, such as automatic enrolments, opt-out policies, and pre-selected options is inconsistent. In general, the use of defaults has received less support than other interventions and support has been much more variable depending on the aim promoted by the intervention. For instance, some studies found support for defaults that promoted pro-social aims. Different studies reported high acceptability of the encouraged and mandatory use of default settings that make large electricity companies automatically switch costumers to green energy providers (Loibl et al., 2018; Reisch & Sunstein, 2016; Sunstein, 2017b). Likewise, some defaults that promoted pro-

self aims were also supported. Jung & Mellers (2016) and Sunstein (2017b) found high acceptability of the promotion of employees' automatic enrolment in pensions-saving plans. Likewise, Jung & Mellers (2016) found that a majority of people supported default settings in social networks to improve privacy; and Sunstein (2017b) found moderate support (approval rating of 53% and 58%) for automatically registering eligible citizens as voters and automatically changing women's last names to that of their husbands when they marry.

However, some defaults were also opposed. When they used a simple description of defaults without including any aims, Tannenbaum et al. (2017) found low levels of support for their use. Likewise, different studies have found that the majority disapprove of the use of defaults that involve economic costs, for instance, when they are used to automatically deduct donations to charity and carbon offsetting fees (Hagman et al., 2015; Jung & Mellers, 2016; Loibl et al., 2018; Reisch & Sunstein, 2016; Sunstein, 2017b). Interestingly, Sunstein, (2017b) found that people opposed defaults that went against personal and social values; for instance, people rejected defaults that automatically registered others in political and religious affiliations as Democrats and Christians. Likewise, he found that people rejected defaults that contradicted social conventions, for instance, automatically changing men's last names to that of their wife's when they marry.

- *Individual and socioeconomic factors*

Studies on attitudes towards nudges have looked at how individual and socioeconomic factors affect people's opinions of nudges. Literature has shown that while some of these factors are relevant, their effect is only moderate, and it did not find particular population groups in favour or against nudges.

In terms of sociodemographic characteristics, very few correlations have been significant. Studies have found that women are more likely to approve of nudges than men both in general and in specific domains (Evers et al., 2018; Hagmann et al., 2018; Loibl et al., 2018; Reisch & Sunstein, 2016; Sunstein et al., 2018). However, other traditionally relevant variables, such as educational attainment, income and occupation, seem to not be related to nudge acceptance. In representative surveys in Australia, Brazil, Canada, Japan, China, Russia, South Africa, and South Korea, Sunstein et al. (2018) found that age had a significant effect and operated differently for different nudges. Older people preferred information campaigns and mandates, while younger people were more likely to approve of more intrusive intervention such as manipulative messages and defaults. Likewise, they found that education had an effect; the number of years people attended school correlated with higher approval ratings for mandates and information and lower approval of subliminal messages. Sunstein et al. (2019) also found correlations from the results of nationally representative surveys in Germany, Denmark, South Korea, the US and Belgium (Flanders). Higher formal education correlated with lower approval of nudges, people living in cities tended to be more supportive than people who lived in villages or the countryside, and the number of children people had correlated positively with approval. However, except for gender,

literature suggests that there is no systematic correlation between sociodemographic characteristics and support for nudges.

Political affiliation and political views do not seem to be a significant determinant in the public's views on nudges. Some studies reported different support for nudge interventions based on political views. Sunstein (2017b, 2017c) found that nudges that are about politically contested ends tended to reflect partisan differences between Republicans and Democrats. Similarly, some studies found that people with left-wing and centrist views, and people with a deep concern for the environment were slightly more in favour of nudges than conservatives and right-wing people (Jung & Mellers, 2016; Reisch & Sunstein, 2016; Sunstein et al., 2019). However, in general, studies reported no or only minor effects of political affiliation and political views in the approval of nudges, indicating that individuals from across the ideological spectrum support nudges (Evers et al., 2018; Reisch & Sunstein, 2016; Sunstein et al., 2018).

Some studies have investigated how individual factors such as individual type of thinking, and level of empathy affect preferences for nudges. In a survey experiment conducted in Sweden and the US, Hagman et al. (2015) explored how individual differences in thinking and feeling influenced attitudes towards nudges. The authors found that individuals with individualistic views were less accepting of nudges than those with communitarian views. Likewise, individuals with analytical thinking were less likely to perceive nudges as intrusive to freedom of choice than individuals with intuitive thinking. Jung & Mellers (2016) found that empathetic people tended to manifest higher support for nudges. By contrast, individualists tended to oppose all types of nudges, and individuals with a strong desire for control rejected intrusive interventions. These results provide interesting insights, but more research is needed to understand further how individual factors impact nudge acceptance.

- *Cross-countries differences*

Beyond individual and sociodemographic characteristics, studies have investigated how attitudes towards nudges differ between countries. Several studies indicate that nudges are approved of in many countries with diverse cultural and political backgrounds. At the same time, distinct levels of approval have emerged in different countries. Previous surveys have defined three categories of countries according to their different degrees of nudges acceptability (Loibl et al., 2018; Reisch & Sunstein, 2016; Sunstein et al., 2018, 2019). Firstly, a majority of countries belong to the “principled pro-nudge nations”, within which a substantial majority approve of nudges, yet they disapprove of some of the interventions. The US, Canada, Australia, a majority of countries in Europe, such as Italy, Germany, France and the UK, plus Brazil, Russia and South Africa fall into this category. Secondly, literature has found nations that are “nudge enthusiasts”, in which the support for all types of nudges is overwhelmingly high. China and South Korea fall into this category. By contrast, some nations fall into the “cautiously pro-nudge” category, in which, while some nudges are accepted, the overall support for

nudges is lower than in other countries. This is the case in Japan, Denmark and Hungary. It is unclear what causes these variations. In the case of Hungary, some studies attribute the results to the low levels of trust in institutions. The results in Japan and Denmark are more difficult to assess.

- *The aim of the nudge*

Several studies indicate that the most critical factor in forming opinions about nudges is the aim of the nudge. People approve of nudges if they support their purpose and disapprove of nudges if they disagree with their goal. In a series of questions about interventions with deliberately controversial aims, Sunstein (2017b) found that a large majority disapproved of all of the nudges. For instance, people typically favoured the use of labels and educational and information campaigns. However, he also found that they disapproved of their use if the aims of the interventions were illicit. Examples of this include when labels provided information about the manufacturing of a product in a country with communist ties or when an education campaign was about convincing new mothers to stop working and care for their children at home.

On the other hand, when people support the aim, they are much more likely to support any nudge type. For instance, even the most intrusive nudges to promote healthy eating are accepted (Evers et al., 2018; Petrescu et al., 2016). Likewise, Reisch & Sunstein (2016) and Sunstein (2017b) found moderate support for the use of subliminal messages to reduce obesity, probably driven by the fact that people agreed with the aim. Similarly, the disparity of results regarding defaults also suggests that supporting or opposing their use is related to how people feel about the aim, not how they feel about the intervention. Hagman et al. (2015) also found that pro-social nudges have significantly lower acceptance compared to pro-self nudges, both in Sweden and in the United States. These results confirm the relevance of the aim and that (dis)approval of the aim implies (dis)approval of the nudge.

Because people's acceptability of nudges is directly related to how they feel about the nudge's aim, their opinion on nudges tends to reflect political views. For instance, political differences in the public opinion of nudges appear when the aim is politically contested (Sunstein, 2017b). Tannenbaum et al. (2017) showed that attitudes towards nudges reflected a *partisan nudge bias*. People agreed with nudges that supported policies that aligned with their political views or that were implemented by politicians whom they supported. However, when the interventions were described without partisan cues, the differences for political affiliation disappeared. Consequently, it seems that when the aim is politically contested, more polarisation is likely to occur.

The importance of the aim of nudges on their acceptability has also been raised by some studies, revealing that people who engage in behaviour targeted by a nudge are less likely to accept its implementation. In that sense, individual preferences and behaviour seem to play an important role. People do not want to be nudged if the nudges target their behaviour. For instance, Sunstein et al. (2019) showed that meat consumption correlated negatively with the approval of a mandate to regulate meat

consumption; smokers disapproved of different interventions to reduce smoking and people who drank alcohol were more likely to disapprove of nudging in general. Similarly, Hagmann et al. (2018) found that support for interventions to lower sugar intake was higher among people who do not or only infrequently engage in the consumption of sugar and lower among those with a higher consumption of sugar-sweetened beverages. Likewise, they found that overweight people tended to be less supportive of public health campaigns, reducing availability, and reducing the sugar content in products compared to individuals who were not overweight. At the same time, the relation between individual preferences and nudges also has a positive side. When nudges pursue aims that align with personal beliefs and personal preferences, they are more likely to be supported. For instance, Hagmann et al. (2018) found that dieters and people with sugar issues and health concerns were much more likely to support interventions to reduce sugar intake. Likewise, they also found that women were more likely to support nudges, especially in the case of healthy eating, probably due to the fact that they tend to be more concerned with their health status than men.

- *The type of nudge*

Another crucial factor in forming opinions about nudges is the type of nudge. People's opinions of different nudges seem to indicate that they intuitively distinguish between nudges with different degrees of intrusion. Accordingly, people tend to prefer nudges that are noticeable and engage deliberation, such as labels and feedback, rather than nudges that work covertly and limit deliberation and attention, such as defaults. Research has made several distinctions when it comes to exploring opinions towards nudges with different degrees of intrusiveness. Some studies have distinguished between System 1 and System 2 nudges (Jung & Mellers, 2016; Sunstein, 2017b). Similarly, Davidai & Shafir (2018) and Petrescu et al. (2016) have described nudges as working consciously and subconsciously, and Felsen et al. (2013) talked about *overt* and conscious nudges vs *covert* and subconscious nudges. There is a shared understanding underlying these distinctions that some nudges tend to engage people's agency more, operate with conscious awareness and are more transparent. By contrast, others tend to work with limited attention and reflection and be more automatic. Below, I will use the distinction of automatic nudges vs reflective nudges to differentiate between these two general types. Interventions described as System 1 nudges, covert nudges and nudges that work subconsciously fit into the automatic category. By contrast, interventions described as System 2 nudges, overt nudges and nudges that work consciously fit into the reflective category.

As discussed in Chapter 2, automatic nudges have been the target of much criticism in the philosophical discussion on the ethical validity of nudges. The essential idea is that automatic nudges undermine human agency and autonomy. Literature on attitudes towards nudges seems to indicate that people also share part of the concerns expressed in the ethical debate. In general, reflective nudges tend to receive greater support than automatic nudges. For instance, a series of surveys found that labels and educational campaigns to promote a healthy diet received more support than interventions that changed

the placement or size of specific products for consumption (Evers et al., 2018; Petrescu et al., 2016). Likewise, while the use of reflective nudges tended to be generally supported, the acceptance of automatic nudges depended more on the aim they promoted.

Some studies have specifically studied people's preferences for both types of nudges. In a contrastive vignette experiment conducted online with samples from the US and Canada, Felsen et al. (2013) compared attitudes towards *overt* nudges that targeted conscious processes and *covert* nudges that targeted subconscious processes. Subjects chose which programme they preferred in different domains (healthy eating, purchasing behaviour, exercise, savings and productivity), either one using an overt nudge or another using a covert nudge. Overall, they found a general preference for overt nudges. The difference between the two nudges was statistically significant, with the exception of the productivity scenario. Likewise, the authors reported some elements that indicated that these preferences were motivated by people's desire to make authentic choices. Firstly, the authors found that people's perception of manipulation correlated with disapproval of nudges. Secondly, the authors found a correlation between how respondents valued the authenticity of their choices in each scenario, and how likely they were to opt-out of the programme.

In a similar study comparing preferences for both types of nudges, Jung & Mellers (2016) also identified a general preference for reflective nudges. Using contrastive vignettes, the authors tested preferences for automatic and reflective nudges in five scenarios: medical coverage, retirement savings, credit card payments, organ donation and water consumption. In all cases, reflective nudges received greater support than automatic nudges. Likewise, only two nudges were opposed: the automatic payments of credit card bills and the automatic registration of people as organ donors, both automatic nudges. The authors also explored the reasons behind these preferences with a series of co-variables and found that automatic nudges were viewed as being more paternalistic and threatening to autonomy than reflective nudges. By contrast, reflective nudges were viewed as being more effective and necessary for making good decisions. Again, the idea that people like to retain their agency is present.

Other studies have also identified a general preference for reflective nudges. Arad & Rubinstein (2018) found a general preference for reflective nudges over automatic nudges in the healthy-eating and retirement savings domains. They found that negative attitudes towards nudges were associated with concerns about manipulation. Their findings suggest that people experience a psychological reactance to being nudged since, in some cases, they choose to act contrary to what the nudge is intended for as a reaction against it. Finally, their study found that the type of nudge was important, as respondents were willing to sacrifice efficiency over accepting the use of a more intrusive intervention when faced with a trade-off between efficiency and nudge type.

Similarly, Niker et al. (2018) also explored how people accepted the effect of rational and arational social influences in their decision-making. In another study using contrastive vignettes in the areas of

voting and healthy eating, respondents were informed that the opinion of a third party had influenced their choice (of whom to vote for or what to eat). In one condition, the third party had influenced their choices by giving them reasons (rational influence). In the other condition, the third-party had not offered any reasons (arational influence). The authors found that rational influence was perceived as less undue; this was less likely to override subjects' control over their behaviour.

Several studies confirm the general preference for reflective nudges over automatic nudges. They suggest that the reason behind this preference is that people have an intrinsic desire to be aware of the factors that influence their choices and like to control their decisions. However, it is still unclear why or under what circumstances people dislike automatic nudges.

Some studies suggest that preference for automatic and reflective nudges may be much more flexible than the literature indicates. Preferences among interventions are not drastically different. For instance, in a survey experiment about intervention to reduce sugar consumption, Petrescu et al. (2016) found no significant effect on the acceptability of nudges as a result of describing them as working via non-conscious or conscious processes. Likewise, research indicates that preferences over automatic and reflective nudges change when other factors are considered. For instance, Sunstein (2017c) showed that preferences over automatic and reflective nudges varied when subjects knew about the intervention's effectiveness. When subjects were informed that automatic nudges were more effective, acceptability for automatic nudges increased significantly. Consequently, while people might preferred reflective nudges to promote people's agency, they also cared about effectiveness and, when this aspect was taken into consideration, they were more likely to support automatic nudges.

Davidai & Shafir (2018) argued that previous research has exaggerated the preference for reflective nudges over automatic nudges. In the majority of studies, people compare and choose between either an automatic or reflective nudge. The authors argued that this method of evaluation makes the differences between the two types of nudges salient, and, when compared, one appears to be more appealing than the other. In a series of survey experiments about decisions regarding health plan coverage, paying bills on time and retirement savings, the authors showed that, in a separate evaluation condition, the acceptability of automatic nudges increased, and that there were no statistical differences with the approval of reflective nudges<sup>16</sup>. The authors also found that the method of evaluation (either jointly or separately) affected people's views on whether the nudge was paternalistic. Automatic nudges again appeared to be less paternalistic when evaluated separately than when evaluated jointly. Finally, the authors also compared support for both types of nudges when people were given information about their effectiveness. They found that the information on policy effectiveness dominated participants' judgments when the policies were evaluated jointly, but not when they were judged separately. The

---

<sup>16</sup> It is worth noting that due to difficulties in describing the two types of nudges, most studies use words such as "conscious", "informing", "explaining", and "educating" to describe reflective nudges and "automatic" and "unconscious" to describe automatic nudges. This wording can further exaggerate preferences for automatic nudges over reflective nudges.

effect of the evaluation method implies that people do not judge the merits of the policy itself but how it compares to the alternatives.

Nudge type is undoubtedly a relevant factor. While research has revealed a general preference for interventions that preserve agency, some studies also indicated that people do not categorically reject automatic nudges. Support for automatic and reflective interventions is malleable. It can be affected by how interventions are presented and by the inclusion of additional information, such as information on effectiveness. Therefore, further exploration on how these factors and others can influence the acceptability of automatic nudges is necessary to see which nudges fit different contexts.

- *The effectiveness of the interventions*

One of the strongest predictors of nudge's acceptability across different studies on attitudes towards nudges seems to be the perceived or reported effectiveness of interventions. Firstly, several studies have shown that people support nudges that they think will be effective. In the absence of real information about the effectiveness of the interventions, people favour those nudges that they expect will be effective (Hagmann et al., 2018; Petrescu et al., 2016). Secondly, when subjects are informed about the effectiveness of the interventions, this information strongly influences their preferences and people support effective interventions. Some studies have shown that information on effectiveness modifies preferences for automatic and reflective nudges. When people are informed that automatic nudges are more effective, support for these interventions increases significantly (Arad & Rubinstein, 2018; Davidai & Shafir, 2018; Sunstein, 2017c).

However, some studies have shown that people have inaccurate beliefs about the effectiveness of interventions. In general, less intrusive interventions tend to be perceived as more effective than more intrusive interventions (Evers et al., 2018; Petrescu et al., 2016; Sunstein, 2017c). These results suggest that effectiveness is a double-edged sword. When people are informed about the effectiveness of interventions, they update their beliefs about interventions and tend to favour more effective interventions. However, when people do not know the real effectiveness of an intervention, their perceived effectiveness seems to be influenced by their values. Sunstein, (2017c) suggested that subjects say that an intervention is effective not because they think it is, but because they like the intervention. Petrescu et al. (2016) also suggested that people's inaccurate beliefs about effectiveness might reflect a process of cognitive consistency by which less accepted/liked interventions are seen as less effective.

- *The context of choice*

Acceptability of nudges varies depending on the context in which they are applied. As argued in Chapter 2, the control that people have over their behaviour in specific domains and how much they value making particular choices are two factors that should be taken into consideration when valuing the

ethical acceptability of nudges. Studies on attitudes towards nudges have confirmed that people's ideas and perceptions of both factors are essential to determining their opinions on nudges.

Nudges are designed on the premise that people's choices are highly influenced by the context of choice and often beyond an individual's control. Consequently, people's subjective perception of these factors may affect their opinion of nudges. Some studies have shown that these factors correlate with nudge acceptance. Gyrd-Hansen & Kjær (2015), found a strong association between people's self-reported behaviour and their perception of self-control in different areas of health-related behaviour. They found that support for interventions to reduce smoking was higher among smokers with lower perceived self-control.

Similarly, in a survey exploring the acceptability of nudges in reducing sugar consumption, Hagmann et al. (2018) found that a higher subjective perception of self-control was associated with a lower approval of interventions. By contrast, a lower perception of self-control was associated with a higher acceptance of nudges. Likewise, Petrescu et al. (2016) also found a positive relationship between participants attributing being overweight to influences in their environment and their support for nudges to reduce the consumption of sugar-sweetened beverages.

Along similar lines, Felsen et al. (2013) discussed how the context might modify the acceptability of nudges. In an experiment about preferences for automatic vs reflective nudges, they found that describing nudges as working either subconsciously or consciously had an effect on some contexts (healthy eating, purchasing behaviour, exercise and savings), but no effect on the acceptability of nudges designed to encourage productivity. The authors argued that the effect of the context might be to do with a subject's perceived capacity to control their own choices in that domain and their knowledge about how individuals decide in particular contexts. Accordingly, Felsen et al. (2013) suggested that in those areas in which individuals experienced less control over their decisions, they were more willing to accept an intervention to change behaviour. They also speculated that in a context in which individuals felt or knew that their choices were usually unconscious, they were more willing to accept a subconscious influence to change their behaviour. Hagan et al. (2015) found that considering an intervention intrusive on individual freedom of choice might not correlate with finding it less acceptable. In their study, a majority of respondents found the use of nudges intrusive on their freedom of choice, yet they still approved of their use. Likewise, they found that people with analytical thinking were less likely to perceive nudges as intrusive on their freedom of choice, which may be explained by the fact that they understand the role that unconscious influences play on decision-making.

Subjective perceptions and beliefs about how others make decisions, how the environment influences their decisions and how much self-control they have can also impact on people's opinion on nudges. Gyrd-Hansen & Kjær (2015) found that support for interventions on health-related behaviour were higher among those not targeted by the intervention. These results might suggest that people understand

that for those that engage in self-harming behaviour, such as smoking, it is sometimes challenging to change their behaviour. Likewise, support for nudges seems to be higher among women and empathetic people, which could also indicate that they support nudges out of worry for others.

Overall, these findings suggest that people that are aware of their own and other people's limitations and understand the effect of the context of choice in decisions are more likely to support nudges interfering in their choices. Exploring the degree to which people perceive that they have retained control over their decision-making in different domains is, therefore, useful when discussing the acceptability of nudges.

How decisions are made in different contexts is an important variable, but not the only relevant one. In some contexts, people might struggle to decide and may even be aware of their limitations and self-control problems, yet they will refuse to delegate their choices. As argued in Chapter 2, when valuing the relevance of the context of choice, it is not only significant how the choice is made, but also, how much people value making that specific choice. For instance, Gyrd-Hansen & Kjær (2015) investigated whether perceived lack of self-control in specific health-related areas (child immunisation, alcohol consumption, dental check-ups, buying healthy food, visiting a general practitioner, smoking, participation in screening programmes, volunteering as an organ donor and exercise) correlated with higher acceptability of interventions that target those areas. The authors found that people experienced low self-control when questioned about exercise and becoming organ donors. A priori, these findings suggest that government intervention in these areas could be justified and receive greater approval by people. However, Gyrd-Hansen & Kjær (2015) found that despite organ donation being an area where respondents experienced self-control problems, respondents disapproved of the interventions to promote organ donation. Looking at studies on attitudes towards nudges as a whole, interventions with different degrees of intrusion intended to improve exercise and healthy eating were broadly accepted. By contrast, the use of the same type of interventions to promote organ donation usually met with disapproval. The contrast between these two cases indicates that the issue at stake is relevant.

Literature on attitudes towards nudges provides some insight that suggests that people care differently about different choices. In general, how people value choices in specific domains is usually approached as an individual issue. As previously considered, the essential idea is that those who have preferences for A will reject nudges that target A. People reject nudges when they disagree with their ends or when they represent an inconvenience to what they want to do. However, consistent disapproval of the same type of interventions in specific choices or domains indicates that the acceptability of some nudges is not only an individual issue but a collective one. For instance, the overall rejection of defaults to promote organ donation might indicate that people think that this is a sensible choice; a choice that should be made with full attention.

Similarly, Sunstein (2017b) suggested that respondents not only reject nudges that are inconsistent with their interests or values but also reject nudges that are against the interest and values of most choosers. He found that people tended to reject automatic nudges that involved economic losses. Indeed, several studies have indicated that the automatic deduction of donations to charity, carbon offsetting fees and other issues related to money are not approved of by the majority (Jung & Mellers, 2016; Loibl et al., 2018; Sunstein, 2017b). Likewise, the automatic designation of political or religious affiliation broadly met with disapproval, even by those who could potentially benefit from it. Finally, when nudges went against social conventions or involve more sensible issues, they were generally opposed (Sunstein, 2017b). Overall, literature indicates that people's opinion of nudges varies depending on the nature of the decision in that domain, and the issue at stake.

- *The nudger*

A final element that seems to be relevant for people when it comes to supporting or rejecting nudges is the entity implementing the nudge. The opinion that people have of the nudger is likely to play a role in their support for those nudges. As mentioned above, Tannenbaum et al. (2017) showed that people approved of nudges when they were implemented by the party or politician they support but opposed nudges when they were applied by their rivals. Likewise, other studies have emphasised that the trustworthiness of the source that implements the nudge is essential, with people supporting nudges that are implemented by a source that they trust. Niker et al. (2018) showed that when a pre-authorized agent influenced choices, the influence was judged as being less intrusive than when it came from a non-authorized agent. Consequently, when agents authorized other agents to influence their behaviour, those nudges were much more likely to be supported. Similarly, Evers et al. (2018), found that people find private companies and experts on the subject at stake to be significantly more trustworthy than policymakers. Accordingly, nudges implemented by experts and industry are more approved of than nudges implemented by policymakers.

Along similar lines, some surveys found a relationship between trust in institutions and support for nudges. In Sunstein et al. (2019), trust in institutions correlated strongly with nudge approval. People who trusted public institutions trusted the nudger and, consequently, would be more likely to support nudges. Likewise, Loibl et al. (2018) and Reisch & Sunstein (2016) suggested that lower approval of nudges in Hungary compared to other countries in Europe was a result of citizens having low trust in the Hungarian public institutions. Overall, the entity that applies the nudge is an issue that may impact on people's opinions on nudges.

### **3. Unresolved questions and future directions**

The previous sections in this chapter have presented the main characteristics and findings of studies on attitudes towards nudges. Section 2.1 summarised the methods and attributes of the studies, with particular emphasis on identifying the most common questions about nudges, and which interventions

and domains tend to appear in these types of studies. Section 2.2 summarised the results of these studies by reviewing the factors that literature has identified as being most relevant to forming opinions on nudges. The review has identified consistent methods and findings, but also has exposed gaps in the literature and identified questions that future empirical research should address.

In general, the results of studies on attitudes towards nudges provide three central themes.

Firstly, literature reveals that acceptability towards nudges is high; people favour the use of different nudges applied to both pro-self and pro-social domains. Studies indicate that people do not categorically oppose nudges and value them as a positive tool for behavioural change. In terms of the general acceptability of nudges, the review also indicates that studies tend to include questions about the same type of nudges and their use in the same areas while leaving out some nudges and relevant domains. In this sense, future studies should include a more diverse catalogue of nudges and more questions about nudges promoting pro-social goals. Adding questions about more diverse interventions and their use in pro-social domains would be beneficial to advancing the research and understand further how people will react to specific nudges.

Secondly, the results on attitudes towards nudges indicate that no particular individual or social factors are especially relevant. While some individual and socioeconomic variables are relevant, their effect is generally small or unimportant. Likewise, while some specific dynamics arise in some countries, a vast majority of nations endorse the use of nudges. Regardless, future studies should continue to explore and control how individual and socioeconomic factors affect the acceptability of nudges. Likewise, countries that plan to implement these types of interventions should collect citizen's views on nudges to ensure their effectiveness and legitimisation.

Thirdly, studies indicate that the acceptability of nudges relates to specific issues in their design and implementation. Supporting the ideas presented in Chapter 2, three of the crucial factors that determine people's preferences for nudges are the aim of the nudge, the type of nudge and the context of choice in which it is applied. Likewise, research indicates that two more factors are important: who applies the nudges and how nudges compared to other policy tools, especially in terms of effectiveness. Future research should explore how these factors affect nudge acceptability.

I shall argue that future research should focus on two different issues to understand the conditions under which nudges receive approval. The first issue relates to the design and implementation of nudges and how the abovementioned factors (aim, type, context, nudger and alternatives) influence nudge acceptability. The second issue is related to how nudges are presented and understood by the subjects and how they can be communicated to the citizens to improve their acceptability. In the two sections below, I discuss future issues relevant to both issues in more detail.

### **3.1. Elements of the design and implementation and nudges acceptability.**

Many questions about how the design and implementation of the nudge affect people's acceptability are still unresolved. Exploring these issues involves addressing questions such as: how do people understand and accept external influences in their choices? How do people value and receive interventions with different degrees of intrusiveness? Which choices and domains are permitted, and which areas are off-limits? Who should apply nudges? How should people be involved in their design?

The first important factor to take into consideration is the aim of the nudge. Research indicates that policy goals are the most relevant factor when it comes to forming opinions about nudges; people have to approve of a nudge's intention to support it. Accordingly, interpreting how people feel about nudges is challenging because the aim of the intervention may determine their opinion. People do not necessarily judge the nudge itself but express an opinion that takes the aim being promoted into consideration. The variety of views towards nudges due to the aims they promote means that if research focuses excessively on looking at how people feel about nudges with different aims, we might end up collecting opinions about policy goals and not about nudges.

I suggest that crucial question to be addressed regarding the effect of the aim on nudge acceptability is what the differences between nudges with pro-self and pro-social goals are. Future research should investigate and address the different acceptability of nudges with pro-self and pro-social goals.

The acceptability of pro-self and pro-social nudges deals with different factors and has different implications. Research on pro-self nudges indicates that people generally accept their use, especially in the areas of healthy eating, exercise and savings. In some cases, however, some pro-self nudges are opposed by people with strong individual preferences for the behaviour that nudges are targeting. In some others, they are rejected by the majority. In the case of pro-self nudges, the acceptance of those targeted by the nudge is crucial. Consequently, if we plan to implement pro-self nudges, we should ensure that the nudges are broadly accepted<sup>17</sup>.

By contrast, the acceptability of pro-social nudges is more challenging. Pro-social nudges go against individual interests to the benefit of social welfare; consequently, many citizens might find their use to be more inconvenient. The acceptability of pro-social nudges has to navigate the trade-off between individual and social wellbeing. In this case, the essential question is how much people can be nudged to promote pro-social goals. To improve the acceptability of pro-self nudges, future studies should consider a number of different issues. Firstly, it would be interesting to study how people react when nudges promote policy goals that have already been justified. Secondly, it would be relevant to study whether nudges are accepted when they improve the design and implementation of programmes that

---

<sup>17</sup> Some pro-self nudges have pro-social implications. For instance, some nudges target individual behaviour that causes negative social externalities. Likewise, in some cases, nudges are useful for the vast majority of people. In this case, if a majority accept nudges, the fact that a specific group rejects their use should not be that important.

are already in operation. Finally, it would be valuable to know how people react when nudges make it easier for people to follow pro-social behaviour. Overall, we need more evidence on how people react to pro-self nudges and how their acceptability differs from support for pro-self nudges.

The second main unresolved question about attitudes towards nudges is how people respond to nudges with different degrees of intrusiveness. As mentioned in previous sections, literature identifies a general preference for less intrusive interventions. At the same time, research suggests that preferences for different types of nudges are malleable and change depending on the context of choice.

Current research on the acceptability of automatic and reflective nudges looks at how people accept these different interventions depending on the policy goal. Research indicates that the policy goal is relevant to shaping preferences for different nudge types; however, it is not fully determinant. People care about policy goals, but they also care about the means that policymakers will employ to achieve them. In some cases, the acceptance of a policy goal is sufficient to accept an automatic nudge; in others, people may agree with the aim but reject the use of intrusive interventions. For instance, in the area of healthy eating, a majority support automatic nudges, whereas, in the case of organ donation, people support the goal but oppose the use of automatic nudges.

I suggest that the acceptability of automatic and reflective nudges is related to the context of choice in which the nudge is applied. Accordingly, future studies should explore the contexts in which people perceive automatic nudges as being permissible or not permissible. As argued in Chapter 2, the context of choice references two aspects, (i) the nature of decisions and (ii) the issue at stake. I argue that the acceptance of intervention with different degrees of intrusiveness depends on people's (subjective and objective) perceptions of both matters.

Firstly, research suggests that the acceptability of automatic nudges might be affected by how particular choices are made in different contexts. In areas in which people's choices are highly influenced by the environment and affected by unconscious and unreflective factors, people might be more accepting of automatic nudges. These issues imply that the acceptability of nudges is related to folk concepts of choice, decision-making and autonomy. In that respect, research also suggests that people have inaccurate beliefs about their agency in their choices. In general, people attribute more self-control to themselves than they have and tend to disregard or be unaware of how many contextual factors affect their decisions. In situations in which this is the case, informing of the effect of the context might be useful. Understanding how people make choices and how they perceive the way in which they make choices in different domains is essential to getting a better understanding of the context in which nudges will be accepted.

Secondly, people's acceptability of different types of nudges is related to the domain in which the nudge is applied. To understand the importance of this area, it is crucial to distinguish it from the aim of the nudge. As argued above, people might agree with the aim of the nudge, but in specific domains, they

may reject the use of intrusive nudges. By contrast, in some cases, people might be indifferent to the aim; however, because the area of intervention is not particularly relevant to them, they are willing to accept intrusive nudges. Future research on the acceptability of nudge should explore which domains are off-limits when it comes to intrusive interventions and in which areas it is acceptable to use more intrusive interventions. Research already indicates that when choices involve economic losses, contradict social conventions and are related to individual religious or political values, people disapprove of the use of intrusive nudges. Future studies should explore further how the area in which nudges are applied affects their acceptability.

Overall, research on attitudes towards nudges has verified that the aim and type of a nudge affect its acceptability. Likewise, some studies suggest that context is important. However, the findings are somehow incomplete because they tend to approach these factors independently from one another when they are, in fact, related. People's acceptability of nudges depends on the interaction between how they feel about the policy goal and how they feel about each nudge in the context in which it is applied. Studies should use more experiments to control how a change in one of these aspects influences people's opinions of nudges in order to understand the ways in which these three factors affect the acceptability of nudges.

Another element of the design and implementation of nudges that is relevant to understanding their acceptability is who applies the nudge. Several studies indicate that trusting the nudger is an essential part of accepting the nudge. To follow this line of research and clarify how this factor affects nudge acceptability, future studies should address three specific questions. Firstly, studies should continue to explore how and why the nudger affects the support of nudges. Along the same lines, literature should explore which nudgers appear to be more trustworthy. Secondly, studies should gather data on trust in governments and public institutions and continue to investigate how these factors affect support for nudges. These studies should also explore whether nudges are supported more when they are applied by regional governments, local administrations or by institutions such as schools, universities and companies. Finally, future research could explore ways of involving citizens in the design and implementation of nudges and the effect that pre-authorising the nudger has on the acceptability of the nudge. The acceptance of nudges should be higher if people are involved in their design and implementation or when they support the nudger. Consequently, a better understanding of these factors is essential.

Finally, the acceptability of nudges varies depending on the alternatives that are available and how they compare to nudges in terms of coercion, intrusiveness and effectiveness. In that sense, support for nudges cannot be studied without exploring how people compare them with other policy tools. As stated, research indicates that people prefer less intrusive nudges. However, it also indicates that people prefer nudges over more coercive measures such as taxes and bans. Policymakers can use different tools

to promote behavioural change, which means it is crucial to continue to investigate the circumstances in which and reasons why people prefer nudges to other interventions.

To sum up, I suggest that literature on attitudes towards nudges should focus on understanding how the different elements of the design and implementation of nudges affect people's support for nudges. In particular, I suggest that future research should address the following questions. Firstly, studies should explore the differences in support for pro-self and pro-social nudges. Secondly, studies should investigate preferences for different types of nudges and explore how these opinions change due to the context of choice in which the nudge is applied. The acceptability of nudges with different degrees of intrusiveness is one of the questions that requires more attention. Addressing these questions means focusing on subjective and folk concepts of autonomy and decision-making, considering how people understand their decision-making process and their capacity to control their behaviour, as well as how much they value choices in general, and choices in specific domains. I suggest that these questions are relevant not only to the debate on the acceptability of nudges but also essential to advancing the philosophical discussion on freedom of choice and autonomy. Thirdly, research should investigate the importance of nudgers. Studies should explore how the acceptability of nudges changes depending on who applies the nudge. Likewise, they should take into consideration whether public engagement in the design and implementation of nudges, and the pre-authorisation of their use affect their acceptability. Finally, support for nudges should include more empirical research on how it compares to the acceptability of other policy tools.

### **3.2. Framing and nudge acceptability**

Research on attitudes towards nudges shows that the acceptance of nudges is malleable and influenced by how people evaluate a nudge. As mentioned in section 2.2., some studies suggest that the acceptability of more intrusive nudges is lower when they are evaluated by comparing them to less intrusive interventions but increases when they are evaluated on their own. Similarly, studies indicate that the acceptability of nudges is higher if they are compared to more coercive interventions such as taxes and mandates. Finally, some studies show that including complementary information about the nudge, for instance, its effectiveness or information about the mechanism by which it works, also shapes people's views of nudges.

The fact that attitudes towards nudges are affected by the method of evaluation and by the inclusion of different information has two important implications for research, particularly for questions about the acceptability of automatic and reflective nudges. Firstly, future studies ought to take into consideration that when collecting opinions about nudges, the way of presenting the information matters. For instance, in studies on the acceptability of automatic and reflective nudges, it seems that the correct approach is to move from a joint evaluation to separate evaluations, or at least to consider how the method influences support for both types of nudges. Secondly, the fact that adding information about nudges

may influence their acceptability implies that how the nudges are framed and presented to people may impact on their opinion. Preferences for nudges, and especially preferences for automatic nudges might depend on how choices are framed and presented to the subjects. Accordingly, I suggest that research should study how to frame nudges to ensure their acceptability.

The judgment of nudges might be affected by automatic processes and heuristics. For instance, when people have to choose between an automatic or reflective nudge, a comparison of the two makes automatic nudges more unlikeable. In the absence of more information about the nudge or the context, people use the salient information to make a quick judgment about the nudge. By contrast, when agents have information about the effectiveness of the intervention, the automatic response favours the more effective intervention. People will again make a quick automatic judgment with the available information. Sunstein (2017c, p. 69) speculates that preferences for automatic nudges are affected by the automatic process and suggests that ‘System 1 favours System 2 nudges and System 2 favours System 1 nudges’. When people have to judge a nudge, a priori, without much knowledge of the nudge and the effect of context of choice, reflective nudges might appear to be more respectful of individual agency. However, when internal and external influences and limitations are taken into account, automatic nudges may gain acceptability.

To explore how people will respond to nudges depending on how they are framed, I suggest that it may be useful to introduce research on ethics and moral decision-making. Accepting or rejecting a nudge is a decision that involves making an ethical judgment. Nudges often involve conflicting goals, for instance, individual wellbeing vs social wellbeing, individual freedom of choice vs social equality, or individual freedom of choice vs individual wellbeing. In a way, accepting or rejecting nudges resembles a moral decision between good and bad interferences, and usually involves trade-offs with different, but equally relevant moral values. For instance, in the case of a pro-social nudge, people might disapprove of it because they care about their immediate interest. However, people also care about other values, such as fairness and reciprocity. Consequently, how we choose to communicate may improve the acceptability of the nudge.

The resemblance between the acceptability of nudges and a moral choice is also relevant for understanding the reasons behind approving or disapproving of different interventions. Moral reasoning is often said to be a result of deliberative judgment based on internal principles on what is right and wrong. From this position, the fact that people disapprove of automatic nudges is understood as a deliberative judgment about how people want their choices to be influenced. However, extensive research on experimental ethics suggests that moral judgments can also be unreflective and automatic, affected by what is salient in the context. People often employ moral heuristics when facing a decision that involves moral issues (for instance, to value whether A is good or bad) (Epley & Tannenbaum, 2017; Sunstein, 2005). When facing competing moral values automatic processes help to resolve the problem but, like other heuristics, they can also result in mistakes (preferring A to B while preferring

B to A). The judgment over automatic and reflective nudges may vary according to how the information is framed and which particular factor is salient.

Future studies should explore whether heuristics and automatic processes shape the acceptability of nudges. In this sense, it would be interesting to use experimental research to test how different elements influence support for nudges. Research should continue to explore the role of information on effectiveness, information about the method of influence, and information about how the context influences choices in specific domains. For instance, some studies suggest that people experience psychological reactance to being nudged because they tend to resist threats to their freedom of choice (e.g., Arad & Rubinstein, 2018; Loewenstein et al., 2015). In this sense, presenting nudges as a correction of how the environment shapes their choices could raise the acceptability of nudges. Studies should also test how people react to nudges under conditions of time pressure and cognitive load or after priming specific stimulus. A better understanding of how people react to nudges and which mechanisms affect their acceptability would improve their design and communication to ensure their acceptability.

#### **4. Concluding remarks**

Studies on public attitudes towards nudges provide interesting results on their acceptability and show that support for nudges is high. The studies also indicate that several factors about the design and implementation of nudges affect people's preferences for nudges. The literature review identifies significant trends and tendencies, but many questions require further empirical research. The following chapters address specific questions about the public's acceptability of nudges. The three studies are intended to contribute to the current discussion on attitudes towards nudges and provide answers to the central theme of this thesis: the debate on the positive and negative implications of using nudges in policymaking.

## Chapter 6

### Public acceptability of nudges in Spain: a nationally representative survey

#### 1. Introduction

The following chapter analyses public attitudes towards nudges in Spain. It presents the results of a nationally representative survey ( $n= 617$ ,  $e=3.88$ ) on how people feel about different nudges applied to pro-self and pro-social domains. As stated in Chapter 5, literature has previously investigated attitudes towards nudges within several countries, including certain European countries. However, there are no studies on attitudes towards nudges in Spain. Chapter 6 fills this gap by assessing people's opinions on different nudges applied to a wide array of policy objectives.

Exploring the acceptability of nudges in Spain is relevant in different ways. Firstly, it contributes to the growing literature on attitudes towards nudges by considering the general acceptability of nudges and exploring how various factors affect people's preferences for nudges. Secondly, the study complements the current cross-country comparisons by studying the Spanish case. Finally, the results are relevant to advancing the use of nudges in Spain. In the Spanish context, nudging has attracted significant interest within academic circles and the approach has also recently attracted the attention of some media and public institutions, especially on the regional scale. However, the implementation of nudges is limited, the national government has not explicitly implemented any nudges, and there are no specific teams dedicated to developing them. As a result, public discussion about nudges is minimal. These conditions provide a unique context in which people's opinions about nudges are unlikely to be conditioned by the public debate about them. At the same time, how people feel about nudges is relevant to guiding how nudges should be implemented in policymaking.

The survey analysed the general acceptability of nudges in Spain. It included questions about 17 interventions, which covered different types of nudges and various policy goals. In addition to the survey, the study includes an extensive questionnaire that collected sociodemographic variables, political views and trust in institutions and several behavioural factors. Literature suggests that these variables help explain differences in the support for nudge interventions.

The study documents the general acceptability of nudges in Spain. The results are similar to the findings in other western democracies and suggest that people support the use of different nudges and their implementation on different policy areas. However, the results also indicate that specific elements of the design and implementation affect their acceptability.

Chapter 6 proceeds as follows. Section 2 presents the research questions and hypotheses that guide the empirical analysis. Section 3 explains the materials and methods used in the study. Section 4 details the

results of the nationally representative survey. Section 5 is dedicated to general discussion and explains the main themes and tendencies regarding the public acceptability of nudges in Spain. The chapter ends with the conclusions.

## **2. Research questions and hypothesis**

The literature on attitudes towards nudges analyses the general support for nudge interventions, and how different factors affect people's preferences for nudges. As extensively discussed in Chapter 5, studies on the topic report consistent findings, but many factors require further analysis. The present study investigates the following four research questions and explores the following hypothesis, building on the earlier studies.

- *Research question 1: Do people like nudges in Spain?*

The primary aim of the present study is to study public attitudes towards nudges in Spain. The first relevant question is whether Spaniards approve of the use of nudges as a policy instrument. As stated in Chapter 5, public support for nudges is high in many countries with different cultural, social and political backgrounds. Likewise, cross-country comparisons indicate that, within Europe, nudges are supported by the majority. Following these results, I expect to find general support for the use of nudges and great support for a number of interventions.

*Hypothesis 1. The majority accept nudges.*

At the same time, the literature indicates that the acceptability of nudges varies depending on their policy goal. People consider the objective of a nudge before approving or disapproving of it. Support for nudges is not homogenous, and it differs considerably depending on the purpose that the interventions promote. Consequently, I expect to find differences in levels of support between nudges with different policy goals.

*Hypothesis 2. Variations in levels of support for nudges depend on the policy goals of nudges.*

- *Research question 2: Which nudges do people support?*

The second aim of the study is to identify which nudges people like. As argued in chapters 2 and 5, elements of the design and implementation of nudges affect people's level of support for them. The type of policy goal that a nudge promotes is an essential element. According to their goal, nudges can be classified into two types: pro-self nudges, whose main objective is to improve individual wellbeing, and pro-social nudges, whose main aim is to correct externalities or public policy problems. Literature indicates that people prefer pro-self interventions and are less in favour of pro-social interventions. In line with these previous results, I expect to find more support for nudges with pro-self goals than for nudges with pro-social goals.

*Hypothesis 3. People prefer pro-self nudges than pro-social nudges.*

The type of nudge is another essential element. As argued in chapters 2 and 5, nudges have different degrees of intrusiveness. Some nudges are less intrusive, tend to be noticeable and involve deliberation, whereas other interventions are more automatic and less engaging with deliberation. Literature on the subject reports a general preference for less invasive (or reflective) nudges over more intrusive (or automatic) nudges. People dislike non-transparent interventions and/or nudges that change behaviour without explicit affirmative consent. Since several studies support these results, I expect to find similar results in Spain.

*Hypothesis 4. People prefer less intrusive (reflective) nudges to more intrusive (automatic).*

- *Research question 3. Which factors predict the acceptability of nudges?*

The third aim of the study is to explore which factors predict the acceptability of nudges. Previous studies reveal that different elements affect people's support for nudges. Firstly, people's perceptions about the nature of the nudge, specifically about how effective and manipulative they perceive the nudge to be, affect their support for nudges. In line with previous results, I expect to find greater support for interventions that people see as being effective and lower support for nudges that people perceive as being manipulative.

*Hypothesis 5. People's subjective perceptions of effectiveness affect their support for nudges. Those that believe that nudges are effective are more likely to support them.*

*Hypothesis 6. People's subjective perceptions of manipulation affect their support for nudges. Those that believe that nudges are manipulative are less likely to support them.*

Secondly, several studies control for the effect of sociodemographic variables. As discussed in Chapter 5, some factors are relevant, but in general, their impact is minimal and tends to be related to the aim of the nudge. In line with prior results, I expect no general effect of sociodemographic variables on nudge acceptability.

*Hypothesis 7. Sociodemographic factors do not affect preferences for nudges in any systematic way.*

Thirdly, previous studies indicate that political attitudes, political affiliation, trust in institutions and preferences regarding government interference in both social and individual wellbeing can affect choices regarding nudges in specific ways. On the one hand, research suggests that political affiliation and ideology are not relevant when it comes to predicting nudge acceptability. The effect of both variables is only relevant when the objective of the nudge is politically contested. On the other hand, several studies suggest that trust in institutions and views on government interferences do help predict nudge acceptability. In line with the findings from earlier studies, I expect to find a minor effect of ideology on nudge support and a strong positive relationship between trust in institutions and support for government interference and the acceptability of nudges.

*Hypothesis 8. Ideology does not predict public attitudes towards nudges.*

*Hypothesis 9. Trust in institutions and acceptability of government interference predict nudge acceptability.*

Fourthly, previous studies on attitudes towards nudges report the relevant effect of behavioural factors. Some studies indicate that people's preferences and behaviour affect their support for nudges. When nudges target or promote aims that align with their choices, nudges are supported. By contrast, when nudges target behaviour that they engage in, nudges are opposed. In line with these previous results, I expect individual preferences and behaviour to predict support for nudges.

*Hypothesis 10. Individuals preferences and behaviour predict support for nudges. People support nudges that align with their preferences and behaviour and disapprove of nudges that contradict them.*

Finally, in chapters 2 and 5, I have discussed the importance of the context of choice in forming opinions about nudges. In line with these arguments, some studies indicate that people's subjective beliefs about how the environment shapes decisions is a relevant factor in predicting nudge acceptability. People who think that the environment influences their choices and those of others are more likely to support nudges. By contrast, those that believe that the context plays a minor role or no critical role are likely to oppose nudges more. In line with previous studies, I expect to find more significant support for nudges from those who believe that their behaviour is primarily shaped by the environment (situationism) than from those that think that the behaviour is mostly intentional (dispositionism).

*Hypothesis 11. Individuals' beliefs about how environments shape their own and others' choices help predict nudge acceptability.*

- *Research question 4. How does the acceptability of nudges in Spain compare with results in other countries?*

There is still not much literature on attitudes towards nudges, yet surveys carried out in different countries have made it possible to distinguish countries with varying degrees of acceptability. As discussed in Chapter 5, Reisch & Sunstein (2016), Loibl et al. (2018) and Sunstein et al. (2018) and Sunstein et al. (2019) classify countries into three categories: "principled pro-nudge nations", "nudge enthusiast nations" and "cautiously pro-nudge nations". The majority of western democracies and European countries belong to the first category. Given Spain's similarities with these countries, I expect it to fit the same type.

*Hypothesis 12. The acceptability of nudges in Spain is similar to the acceptability found in "principled pro-nudge nations".*

### 3. Materials and methods

The study employed a survey ( $n= 617$ ) to form a general impression of how people in Spain feel about nudges. To ensure the quality of the data, I used a nationally representative survey by age, gender and region with a marginal error of 3.88%. The recruitment, sampling and survey were performed by Netquest, an international market research company with an IOS certified panel and with whom I worked to monitor each step of the planning, sampling and implementation of the survey. Respondents conducted the survey online using the *Nicequest* app, which works on different devices, including tablets, computers and smartphones. The survey was conducted in May of 2019. Following Netquest's regular compensation practices, participants were rewarded with Netquest points (*Korus*) that they could later exchange within the same app for gifts or donate to a charity of their choice.

- *The survey*

The core of the survey was a questionnaire about 17 different interventions. The nudges evaluated in the study are listed in Table 7. Following other surveys on attitudes towards nudges, the survey included questions about different types of nudges (both automatic and reflective), and nudges applied to a wide array of policy domains (both pro-self and pro-social). The items were selected to ensure the survey gathered sufficient data to be able to form a good understanding of general attitudes towards nudge interventions and analyse how people feel about the use of nudges in different policy domains, which type of nudges people prefer and the relationship between the type of nudge and policy domain. As detailed in Table 7, the survey asked participants about the use of nudges in six pro-social domains and five pro-self domains. Of the 17 items, eight were categorised as automatic nudges and nine as reflective nudges (see the full survey in Table 15, Appendix A).

- *Response variables*

Following the description of each measure, participants were asked: 'Do you agree with the implementation of this measure?'; 'Do you think this measure will be effective?' and 'Do you think that this measure manipulates behaviour? Items were rated on a 5-point-Likert scale ranging from 1 ('Not at all') to 5 ('Yes, totally').

**Table 7. List of 17 nudges, categorised by type of nudge and aim of nudge**

Nudges	Type	Aim
Default registration of citizens as organ donors (subject to opt-out).	Automatic	Pro-social
Public campaign to promote organ donation, emphasising that Spain is the European leader in transplants and encouraging people to sign up as donors.	Reflective	
Default setting in hotels to not clean the towels every day (subject to opt-out).	Automatic	Pro-social
Social comparison of energy consumption between households.	Reflective	
Default selection of donations to charity and third sector on tax forms.	Automatic	Pro-social
Educational campaign to promote charity donations.	Reflective	
Google maps add-in that informs users of the carbon footprint of each transportation option.	Reflective	Pro-social
Optical illusions on roads to control vehicle speeding.	Automatic	Pro-self
Educational campaign that refers to the damage caused by and victims of drivers who exceed the speed limit.	Reflective	
Public campaign with moral and social norms to promote tax compliance.	Reflective	Pro-social
Default food layouts in restaurants and supermarkets to promote healthy choices.	Automatic	Pro-self
Traffic light labels on food to promote healthy choices.	Reflective	
Automatic enrolment of eligible individuals into social programmes and benefits that they qualify for.	Automatic	Pro-self
Gambling control application in which people self-select the time of use and budget limit for each week; once set, it can only be modified at the beginning of the following week.	Automatic	Pro-self
Public campaign to educate the population about the problems associated with gambling, highlighting aspects such as the probability of losing money and the psychological and social problems associated with addiction.	Reflective	
Regulation of alcohol container sizes to reduce alcohol consumption	Automatic	Pro-self
Voting reminders to increase voter turnout	Reflective	Pro-self

- *Co-variables questionnaire*

In addition to the main survey, the study included an extensive questionnaire on complementary predictors of acceptability. Participants answered the questionnaire once they had completed the survey. The questionnaire collected sociodemographic variables, political values and information on individual behaviour and individual dispositions.

Firstly, in terms of sociodemographic characteristics, the sample was representative in terms of age, gender and region. Also, the questionnaire collected information on activity, education, nationality and

religious values. Educational level was measured in 8 categories (no education, first stage school certificate (education up to the age of 10), second stage school certificate (education up to the age of 14), lower secondary school education, upper secondary school education, university education (1st level, three-year courses), university education (2nd level, 4-5 year courses), university education (3rd level - master's degree and PhD). Activity was measured in 7 categories (full time and part-time jobs, retired, unemployed, unemployed looking for the first job, student and housewife/house husband). Nationality was measured using three categories (Spanish / Spanish and others / Others) and religion was also measured using three categories (Catholic/Christian / Other religions / Agnostic or atheist).

Secondly, the questionnaire asked respondents about their political orientation, trust in institutions and attitudes towards government interfering for the social and individual good. Using the Spanish Centre for Sociological Research (CIS) methodology, ideology was measured by asking respondents to self-identify on a scale from 0 to 10 where 0 corresponded to the extreme left and 10 to the extreme right. Respondents were also asked if they trust public institutions and whether they favoured government interference in individual choices for improving personal and social wellbeing; they answered these questions using a 5-point Likert scale ranging from 1 ('Not at all') to 5 ('Yes, totally').

Finally, the questionnaire also included two questions about individual beliefs on the factors that drive behaviour and several questions about individual behaviour and personal preferences. Firstly, respondents were asked two separate questions about whether they perceived their behaviour and the behaviour of others as mostly intentional or mostly driven by the environment. They answered using a 5-point Likert ranging from 1 ('my behaviour and the behaviour of others is largely driven by choices and dispositions') and 5 ('my behaviour and the behaviour of others is largely driven by the context and circumstances that surround me/them). Secondly, the questionnaire included several questions about individuals' behaviour related to nudges. The questionnaire identified a number of relevant factors given the 17 items included in the study. The behavioural question collected the following data: whether respondents were organ donors, their subjective concerns about the environment, whether they are used to making charity donations, their main mode of transportation, whether they have ever had a traffic fine, whether they received or have received social benefits, whether they do their own tax declarations, their gambling habits, whether they voted in elections, their dietary regimes, whether they feel they need help to stick to a healthy diet, and their weekly alcohol consumption.

All of the variables in the post-survey questionnaire were incorporated into the analysis as potential predictors of acceptability (see the full questionnaire in Table 15, Appendix A).

- *Sampling*

Table 8 shows the distribution of respondents for the sociodemographic variables. The sample is a representative sample in terms of age, gender and region. The sample is better educated than the general Spanish population, something that tends to be frequent in samples from online panels. Likewise, people

with full time jobs and Spaniards are overrepresented: about 50% of the sample is working full time, and 97% of the sample has Spanish nationality.

**Table 8. Sample sociodemographic characteristics**

<b>Factor</b>	<b>Distribution of respondents</b>
Age	Range:18-75 Median: 44 Mean + SD: 45.11, 15.3
Gender	Male: 49% Female: 51%
Education level	Primary education or lower: 11% Secondary education: 38% College or university degree (2): 40% Post-graduate degree (master's or PhD): 11%
Activity	Full-time job: 50% Part-time job: 10% Retired: 16% Unemployed and unemployed (first job):8.9% Student: 10% Housewife/house husband: 4%

Some respondents were excluded due to inattentive response. To identify possible inattentive respondents, *Netquest* included an attention check within the survey and excluded those who failed to answer it correctly. However, I also analysed time of response to further ensure the quality of the data. On average subjects completed the survey in 14.70 minutes (Md= 12.67 and SD=7.68). Participants could dip in and out of the survey, so the data has several outliers above the mean. Outliers above the mean are unproblematic because the items in the survey are independent of each other, and respondents were not allowed to leave the survey in between reading and answering questions about each intervention. Three cases were excluded because they completed the survey in under 4 minutes, insufficient time to read the instructions and questions correctly.

- *Statistical analysis*

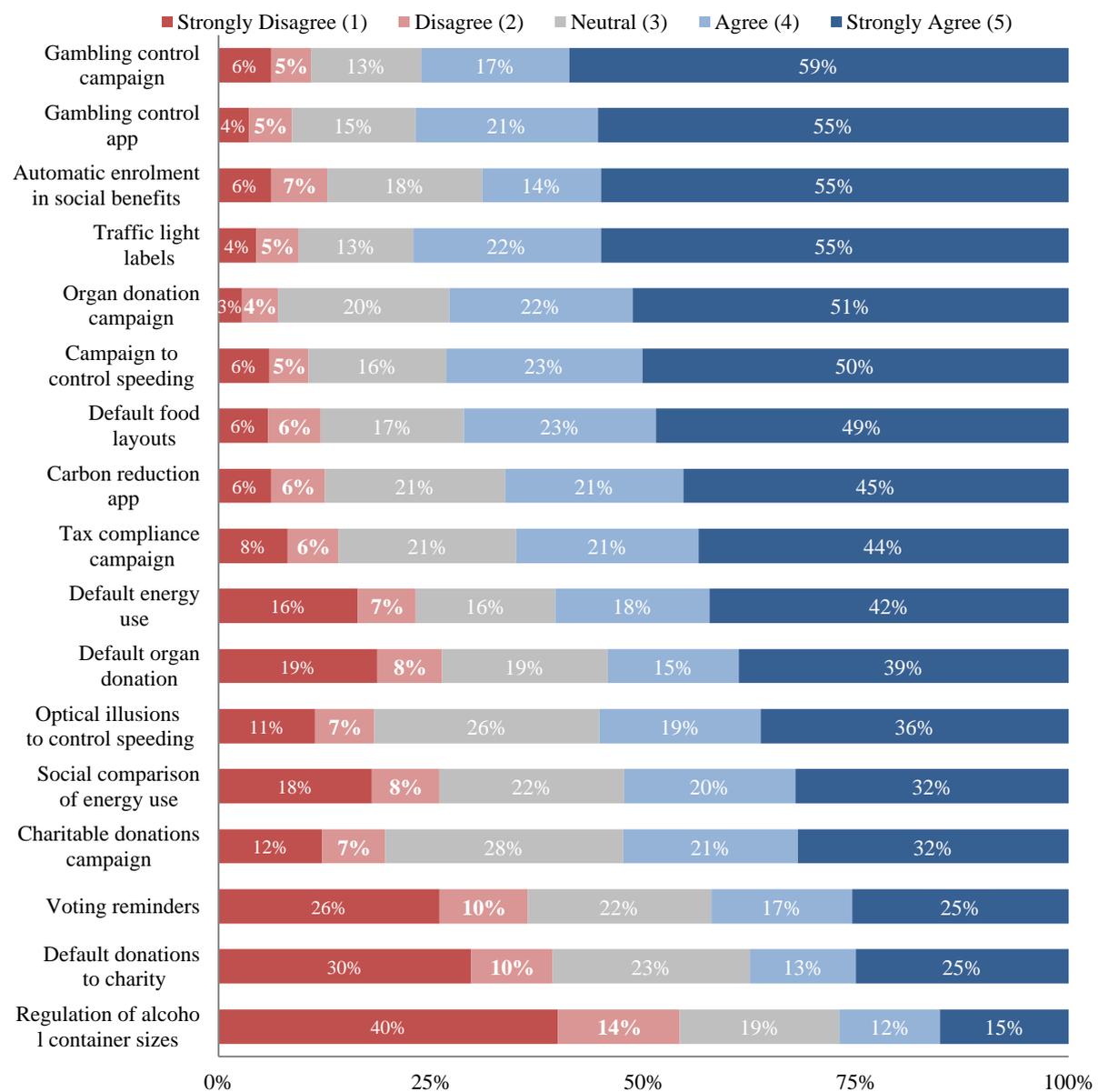
The statistical analysis used different methodologies to ensure a good understanding of the data. The dependent variables in the analysis were Likert-type responses and, as such, they were primarily analysed as categorical variables. The data allowed for multilevel analysis. The relationship between variables was first analysed with chi-square test and then with logistic regression. Multicollinearity was controlled using the variance inflation factor. Responses were grouped for each category in order to analyse the differences between pro-self and pro-social and automatic and reflective nudges. Nudge clusters were treated as numerical variables and the differences between groups were analysed using paired t-tests. Coding and analysis were performed using R.

## 4. Results

### 4.1. Do people like nudges in Spain?

The first meaningful result of the survey identified that support for the nudges presented was high. Figure 1 shows the responses from the survey for the 17 nudges, giving the percentage for each category, with the interventions ordered from most to least supported. As shown, most nudges were supported by the majority of the people surveyed. Fourteen of the seventeen items analysed were approved of, with acceptability ratings ranging from 52% to up to 77% (when ‘strongly agree’ and ‘agree’ responses are combined). The results support hypothesis 1 and are similar to approval ratings reported in other countries with general support for the implementation of nudges.

**Figure 1. Acceptability of the 17 nudges by response.**



Percentages for each category of the five-point Likert scale. Nudges ranked by 'Strongly Agree' responses, from most to least accepted.

The most popular nudges were the two interventions to reduce gambling, the two interventions to promote healthy eating and the national campaign to promote organ donation; all five interventions had support ratings above 70%. Interventions to promote healthy eating were highly supported: 77% of people supported the use of labels with the traffic light system, and 72% supported default food layouts in restaurants and supermarkets. Nudges to reduce gambling were also highly approved of 76% of respondents were in favour the use of an app to self-control gambling, and 77% were in favour of a campaign to reduce gambling that highlighted the risks of addiction. Likewise, 73% of respondents approved of a campaign to promote organ donation by emphasising national success and commitment to organ transplants. The five most supported nudges are not polarising. In all cases, opposition was low, and respondents that felt neutral about the interventions outnumbered respondents who opposed nudges.

Seven out of the seventeen interventions were also supported by the majority, but with more modest approval ratings. Support was still high for the automatic allocation of social benefits to eligible citizens (69% approval rating), for the Google maps add-in app that reports carbon footprints (66% approval rating) and for the campaign with moral and social norms to promote tax compliance (65% approval rating). Opposition was low in all these cases.

However, support was lower for the remaining five interventions that people approved of. About 60% of respondents approved of the default setting in hotels for not cleaning towels every day, and 55% of respondents approved of the use of optical illusions on roads to control vehicle speeding. About 54% of respondents supported the automatic registration of citizens as organ donors; 52% approved of the implementation of a public campaign to promote charitable donations, and 52% approved of the social comparison of energy consumption. In addition to having lower approval ratings, the implementation of some of these interventions was more controversial. As shown in Figure 1, 23% of respondents opposed the intervention related to default settings in hotels to reduce energy consumption, outnumbering those who were either neutral (16%) or only agreed (18%), while only 42% agreed strongly. This is also the case for the social comparison of energy consumption and the automatic registration of citizens as organ donors. In both cases, 26% of the respondents disapproved of the interventions, outnumbering those that felt neutral about implementing them.

Three nudges were opposed: the use of reminders to increase voter turnout, the automatic deduction of donations to charity on tax forms, and the reduction in the size of alcoholic beverage containers. There was considerable disagreement on the reduction in the size of alcoholic beverage containers, with 54% of respondents against the nudge. In the other two cases, disagreement was less forthright: 37% opposed default donations, and 39% opposed the use of reminders to increase voter turnout. Neither of these two nudges received the approval of the majority.

A first look at acceptability data confirms hypothesis 1 and hypothesis 2. Firstly, nudge acceptability is high; people do not categorically oppose nudges and respondents favour different types of nudges applied to a wide array of policy domains. Secondly, support for nudges is not homogenous, and it differs noticeably depending on the aims that the different interventions promote.

As found in previous studies, support for nudges depended on the goal that the nudge promoted. The results of the survey also reflected this theme. The difference in support for interventions seemed to be related to how people felt about the legitimacy of the different aims. For instance, the results indicated that people particularly favoured interventions to promote healthy eating, and interventions to reduce gambling problems. In both cases, people agreed with the use of automatic and reflective nudges that promoted these aims. At the same time, the two nudges aimed at reducing energy consumption had a very similar distribution of responses, again indicating that people valued the interventions according to the aim they promote.

Similarly, I suspect that the negative attitudes towards the three interventions that people disapproved of were related to the types of interventions and their goals. The opposition to the default donations and reduction in the size of alcohol containers is not surprising. Firstly, previous studies have found low support for the use of defaults to increase donations to charities, and for nudges that involve losses, particularly economic losses, without explicit consent (Loibl et al., 2018; Sunstein, 2017b). In some senses, the nudge to reduce alcohol consumption can also be interpreted as a nudge that involves losses if people perceive that a price reduction does not accompany the reduction in size. Secondly, these two interventions are automatic nudges, which, literature suggests, tend to be less supported. Finally, people might perceive these areas as domains in which governments should not interfere. This may also explain the opposition towards the use of reminders to increase voter turnout. Previous studies have found moderate to high support for the use of nudges to increase voter turnout in the USA (Jung & Mellers, 2016a; Sunstein, 2017b, 2017c). However, the voting process there has more barriers than in Spain, where, voter registration is not needed, for example. Accordingly, in the Spanish context, government interference in voting-related decisions might not be seen as a facilitator. Instead, it might trigger concerns of excessive and inappropriate involvement in people's personal decisions on whether or not to vote. In essence, the analysis of the responses revealed that, in general, support for nudges was high and varied depending on the aims of the nudges.

In terms of perceived effectiveness and perceived manipulation, respondents had a less clear opinion, which can be seen from the higher selection of the neutral option on the 5-points Likert scale (see data on perceived effectiveness and perceived manipulation in Figures 7 and 8, Appendix A). The intervention that people perceived as most effective was the automatic enrolment of citizens eligible for social benefits. 40% of respondents 'strongly agree' and 20% of respondents 'agree' that the intervention would be effective. Just below this, respondents believed that the automatic enrolment of citizens as organ donors and the use of a default setting to reduce energy consumption in hotels would

also be effective (56% and 52% respectively, when the ‘strongly agree’ and ‘agree’ responses are combined). The results indicate that people recognise the power of defaults to change behaviour. Likewise, respondents also saw the use of traffic light labels on food and the campaign to promote organ donation as effective interventions (56% and 53% respectively, when the ‘strongly agree’ and ‘agree’ responses are combined).

Overwhelmingly, respondents believed that the most ineffective interventions would be the regulation of alcohol container sizes; 47% of respondents ‘strongly disagree’ and 21% ‘disagree’ with this intervention being effective. Respondents also believed that the campaign with social and moral norms to promote tax compliance and the reminders to increase voter turnout would also be ineffective (55% and 53% respectively, when the ‘strongly disagree’ and ‘disagree’ responses are combined).

Interestingly, the relationship between perceived effectiveness and acceptability was inconsistent. The intervention to regulate the size of alcohol containers and the use of reminders to increase voter turnout were two interventions that were perceived as being most ineffective and also two of the most opposed interventions. However, while the tax compliance campaign using social and moral norms was considered to be ineffective, it was generally supported. Also, the automatic enrolment for social benefits was perceived as an effective intervention and was also highly supported, whereas the use of defaults to reduce energy consumption in hotels and to promote organ donation were perceived as effective yet had modest acceptability levels. For the remaining interventions, neutral responses were between 29% and 42%, thus indicating that respondents were unsure about how effective those interventions would be.

Similar to the results on perceived effectiveness, perceived manipulation had many neutral answers. The results indicate that people did not perceive nudges as intrinsically manipulative. The use of defaults to promote charitable donations was the only intervention perceived as manipulative by the majority of respondents (53%). The automatic enrolment of citizens as organ donors and the limitation on the size of alcohol containers were seen as manipulative interventions by 47% and 38% of respondents respectively (when the ‘strongly agree’ and ‘agree’ responses are combined). The three interventions that people perceived as being most manipulative had lower levels of acceptability. By contrast, people perceived the campaign using social and moral norms to promote tax compliance and the automatic enrolment for social benefits as the least manipulative interventions. In terms of the perceived manipulation of the remaining nudges, there was a high level of neutral responses, suggesting that respondents did not have a clear idea of whether nudges were manipulative or not.

Interestingly, the automatic enrolment of citizens into different outcomes had contradicting perceptions of manipulation. Enrolling those who are eligible into a social benefit was not perceived as a manipulative nudge. By contrast, many believed that enrolling people into making donations to charity or into becoming organ donors was indeed manipulative. The comparison between these interventions

is curious because it shows that people's perceptions of manipulation are not only about the type of intervention but are also related to the intention and outcome of the intervention.

Overall, the results on acceptability, perceived effectiveness and manipulation indicate that people in Spain support the use of nudges as policy instruments. The support for nudges is not unconditional, and the acceptability of their use depends on whether people consider their aim to be legitimate. The different interventions have variable levels of support, and people oppose some nudges. The results also suggest that people perceived nudges as effective and non-manipulative interventions. However, respondents seem to have a somewhat unclear perception of both factors. The acceptability ratings indicate that people assimilate nudges to other policy instruments. However, a more thorough exploration of the data reveals essential themes that are specific to the use of nudges.

## **4.2. Which nudges do people support?**

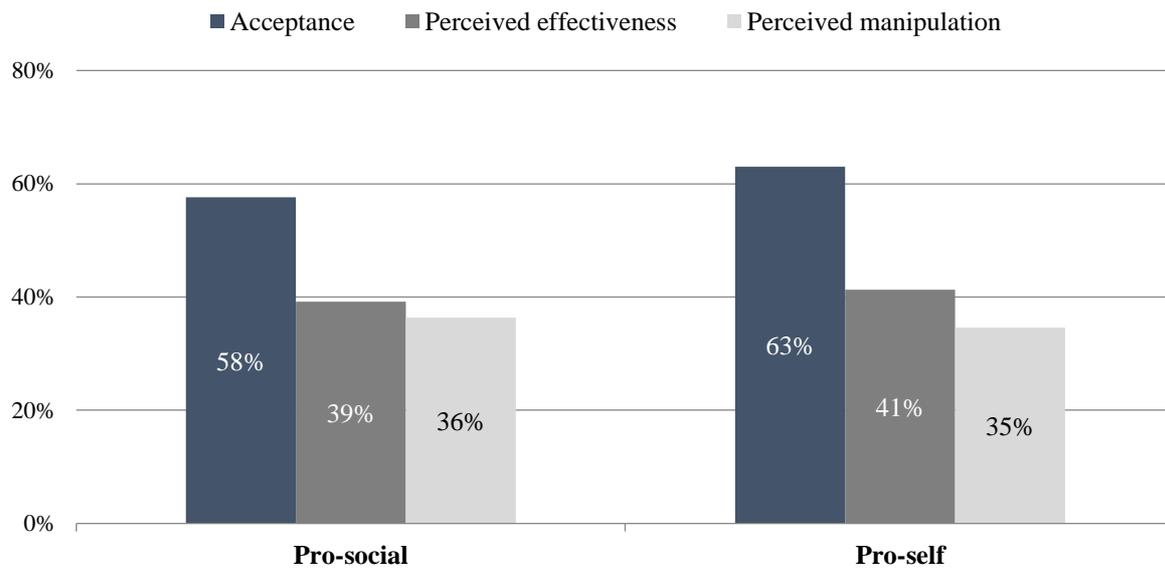
In chapters 2 and 5, I have argued that the type of nudge and the type of policy objective that it promotes are essential factors in forming opinions about nudge interventions. In terms of the policy goal, previous literature reveals that people prefer nudges that encourage pro-self goals over nudges that promote pro-social goals (Hagman et al., 2015). In terms of the types of nudges, previous literature suggests that people prefer less intrusive interventions (reflective) over more intrusive nudges (automatic) (Arad & Rubinstein, 2018; Felsen et al., 2013; Jung & Mellers, 2016; Sunstein, 2017c). To test the effect of both factors, I analysed the differences between the items in the survey that were *a priori* classified as pro-self and pro-social and as automatic and reflective (see Table 7 for the classification).

### *4.2.1. Type of policy goal*

The results indicate that pro-self nudges are preferred slightly more than pro-social nudges. Figure 2 shows the average acceptability rate of pro-self vs pro-social nudges and the average perception of effectiveness and manipulation for both categories. When the acceptance responses were combined, the average acceptability rate of pro-self nudges was 63%, slightly higher than the average acceptability rate for pro-social nudges at 58%. The average perceived effectiveness and the average perceived manipulation of both types of factors were almost identical, indicating that the difference in acceptability between the two types was mostly related to the type of policy goal.

In general, pro-social nudges have more modest approval ratings, with three interventions with approval ratings of around 50%, three around 60% and only one above 70%. By contrast, five of the pro-social interventions have approval ratings above 70%, another was favoured by 69% of respondents, and another was approved of by 55% of the respondents. Interestingly, however, only one of the pro-social interventions was opposed (40% of respondents opposed the default selection of donations to charity on tax forms). By contrast, two pro-self nudges were disapproved of by the majority (54% opposed the regulation of alcohol container sizes, and 36% oppose the use of reminders to increase voter turnout).

**Figure 2. Pro-self vs Pro-Social nudges**



Average acceptability and average perceived effectiveness and perceived manipulation of pro-social vs pro-self nudges, when 'strongly agree' and 'agree' responses are combined. See Table 7 for the classification of nudges between pro-self and pro-social.

To complement the analysis of the differences between pro-self and pro-social nudges, I combined the eight pro-social nudges and the nine pro-self nudges and compared the mean scores for acceptability. I found statistically significant differences between the average acceptability of pro-self nudges (Mean= 3.88, SD=0.76) and pro-social nudges (Mean=3.625, SD=0.83). The difference was significant even though the pro-self category included two of the most disliked nudges (paired t-test difference  $t = 6.0087$ ,  $df = 615$ ,  $p\text{-value} = 0.000000003204$ ). I found no statistical differences in terms of perceived effectiveness and perceived manipulation.

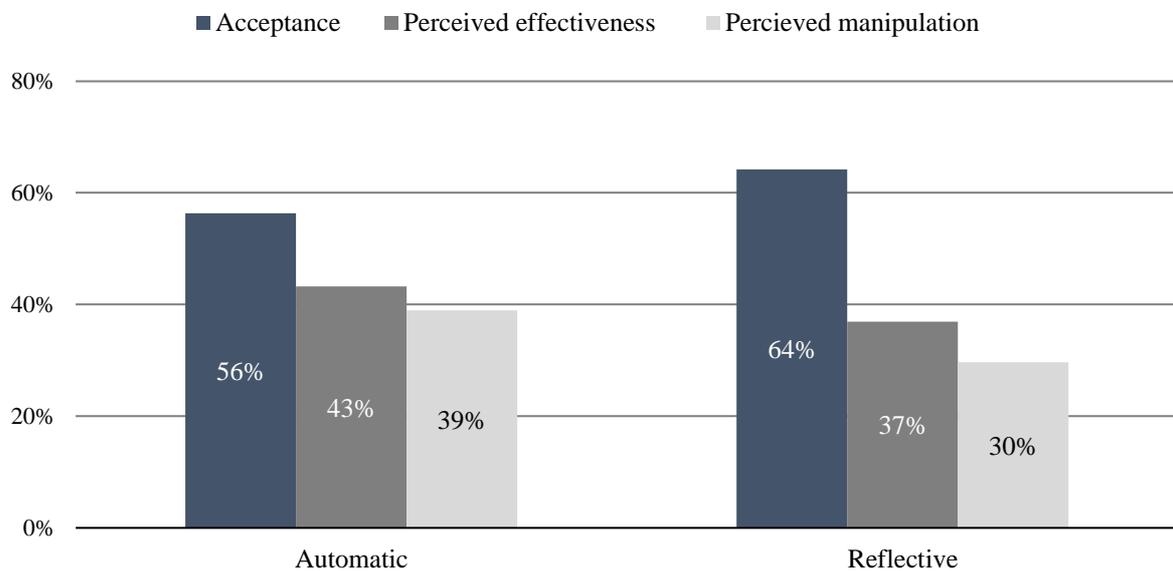
The results of the survey indicate that the type of policy goal is relevant. The findings are consistent with previous studies and confirm hypothesis 3; people prefer pro-self nudges to pro-social nudges. People favour interventions that promote individual wellbeing over interventions that promote social wellbeing. However, the acceptability of pro-self interventions cannot be guaranteed; if people perceive that a pro-self nudge has an illegitimate aim, they will also oppose it. For instance, the regulation of alcohol container sizes and the use of reminders to increase voter turnout were interventions that, while being pro-self, were disapproved of by the majority. The results suggest that a public administration's understanding of which interventions promote people wellbeing might be different from what people understand as positive for themselves and others.

#### 4.2.2. Type of nudge

The results of the survey indicate that the type of nudge is a relevant factor in forming an opinion about nudges. Figure 3 shows the average acceptability rating and the average perceived effectiveness and manipulation of both automatic and reflective types of nudges. When the acceptance responses are combined, the average acceptability rating of reflective nudges is 64%, which is higher than the average

acceptability rating for automatic nudges at 56%. Likewise, automatic nudges are perceived as being more effective and more manipulative than reflective nudges.

**Figure 3. Automatic vs Reflective nudges**




---

Average acceptability and average perceived effectiveness and manipulation of automatic and reflective nudges combining 'strongly agree' and 'agree' responses. See Table 7 for the classification of nudges between automatic and reflective.

---

The differences in acceptability between the two types of nudges seem to be related to people's perception of manipulation. Automatic nudges tend to be perceived as more manipulative than reflective nudges. Accordingly, the results indicate that the theoretical classification of nudges between automatic and reflective matches the respondents' perceptions of the different interventions and affects their acceptability of the different nudges.

In general, all reflective nudges were favoured. Looking at specific interventions, the results are consistent with previous studies and confirm the people favoured less intrusive interventions. The use of educational and public campaigns had a high acceptability rating. The campaigns to promote organ donation, reduce speeding, prevent gambling addictions and promote tax compliance received approval ratings of around 70% and higher. Similarly, the two informative interventions (the use of traffic light labels on food packaging and the Google maps add-in showing information about the carbon footprint of travel options) were also supported, with approval ratings of 77% and 66% respectively.

The reflective nudges that received the lowest approval ratings were the campaign to promote donations to charity (52%), the use of social comparison to reduce energy consumption in households (52%) and the use of reminders to increase voter turnout (42%). The data suggest that the less positive attitudes towards these interventions are a result of a number of different aspects.

Concerning the campaign to promote charitable donations, and the use of reminders to increase voter turnout, the lowest acceptance ratings seem to be related to their aims. In comparison with the other

nudges, respondents did not perceive these two interventions as being particularly manipulative. 31% of respondents believed the use of reminders to increase voter turnout was manipulative, but 39% thought it was not. Likewise, regarding the campaign to promote donations to charity, 31% believed that the intervention was manipulative, but 35% thought it was not. As previously discussed, people may dislike the government interfering in voting decisions and may perceive a campaign to promote donations to external charity organisations as an illegitimate intervention. Overall, the data on acceptability and perceived manipulation suggest that people disagree with the aims rather than with the interventions themselves.

By contrast, people seem to be slightly more concerned about using social comparison to reduce energy consumption; 37% of respondents believed that the intervention was manipulative, and 30% believed that it was not. The respondents' perception of this measure is interesting. In the *a priori* classification, I considered this intervention to be a reflective nudge because it is easy to identify, requires conscious deliberation and does not work if subjects do not perceive or fail to understand its intention. However, people's subjective perception is different; and people seem a bit more concerned about their use. Respondents' negative attitudes towards the intervention might be related to two aspects. Firstly, the intervention is supposed to use data about home energy consumption, and respondents might perceive this as a threat to their privacy. Secondly, people might have a negative perception about being influenced by social comparison. Previous studies have found that people perceive non-rational social influence as being intrusive on their choices (Niker et al., 2018). The fact that 37% of respondents considered the use of social comparison to be a manipulative intervention is compatible with the idea that people reject non-rational forms of social influence.

The lower acceptability of automatic nudges and the fact that people perceived them as more manipulative than reflective nudges confirmed that people have more negative attitudes towards nudges that they perceive as intrusive. In general, the use of automatic nudges was favoured less and more dependent on the aim of the nudge. For instance, the acceptability of default choices was highly dependent on the aim that they promoted. While 69% favoured the automatic enrolment of citizens eligible for social benefits, only 54% supported the automatic registration of citizens as organ donors, and only 38% approved the default selection of donations to charity on tax forms. Similarly, the approval of interventions that modified aspects of their physical environment was also heavily dependent on the aim of the nudge. While 71% of respondents accepted default food layouts in restaurants and supermarkets to promote healthy choices, 54% oppose reductions in the size of alcohol containers.

In some cases, the results suggested that when people favoured the aim, the type of nudge was irrelevant. For instance, default food layouts in supermarkets and restaurants and the use of self-control strategies to control gambling had approval ratings that strongly resemble the acceptance of two of the reflective nudges with the same aims. Interestingly, respondents perceived the two interventions as manipulative,

38% believed that gambling self-regulatory strategies were manipulative, and 45% believed that default food layouts were also manipulative. Yet, when they agreed with the aim, this was less relevant. Likewise, the automatic enrolment of citizens eligible for social benefits is well supported. Consequently, when people think that the aim is legitimate, they support the use of automatic nudges.

However, the results also indicate that people might agree with a policy aim but disagree with the use of an intrusive nudge to promote it. The effect of the type of nudge is salient in two domains. The implementation of a public campaign to promote organ donation received great support and was one of the interventions with the lowest opposition (69% support and 7% opposition). By contrast, the automatic registration of people as organ donors had significantly lower approval rating and many more people were opposed to its implementation (54% support and 27% opposition). Likewise, the public campaign to reduce speeding was well received (73% support and 11% opposition). By contrast, the use of optical illusions with the same aim was less popular and more controversial (55% approval and 18% opposition). Respondents perceived these two interventions as manipulative. The automatic registration of citizens as organ donors was the intervention with the highest score in terms of perceived manipulation. 47% of respondents thought that this nudge manipulated people's behaviour. As found in previous studies, and as discussed in Chapters 2 and 5, the acceptability of automatic nudges does not only depend on the aims they promote but on the context in which they are applied: people accept automatic nudges when they agree with the aims or are unbothered by the decision and disapprove of them in areas in which they like to retain control over their choices and behaviour.

To complement the analysis of the differences between automatic and reflective nudges, I combined the responses from the eight automatic nudges and the nine reflective nudges and compared the mean scores. On average, the mean score for the acceptability of reflective nudges was 3.81 (SD=0.78), which was higher than for the acceptability of automatic nudges at 3.56 (SD=0.80). The difference in acceptability is statistically significant (paired t-test:  $t = -11.154$ ,  $df = 615$ ,  $p\text{-value} < 2.2e-16$ ). In addition, the difference between the two types of nudges was also significant in terms of perceived effectiveness and perceived manipulation. On average, automatic nudges were perceived as being both more effective and more manipulative than reflective nudges. (Perceived effectiveness, automatic: Mean=3.25, SD=0.71 vs reflective: Mean=3.11 SD= 0.74;  $p\text{-value}= 0.000002122$  // Perceived manipulation, automatic: Mean=3.12, SD=0.80 vs reflective Mean=2.88, SD=0.86,  $p\text{-value}= < 2.2e-16$ ).

Overall, the results confirm hypothesis 4; people preferred reflective nudges over automatic nudges. The preference for reflective nudges over automatic nudges appeared to be related to concerns about intrusiveness, as people perceived automatic nudges as being more effective in changing behaviour and more manipulative than reflective nudges. However, people were not categorically opposed to automatic nudges; instead, they supported them in specific contexts and if they were used to promote a legitimate aim. Respondents supported automatic nudges that promoted healthy eating, that improved

the running of social benefit programmes and that controlled gambling. By contrast, rejection of automatic nudges was higher in specific contexts in which people appeared to prefer to retain control, such as alcohol consumption and driving, and in contexts that may have more meaning, such as organ donation.

### **4.3. Which factors predict the acceptability of nudges?**

To investigate the influence of covariables on attitudes towards nudges, I explored the relationship between different variables and conducted a logistic regression analysis for each of the nudges included in the survey. The dependent variable was nudge acceptability. Firstly, I explored the relationship between variables using contingency tables and the chi-square test of independence. Secondly, I conducted logistic regressions considering the variables that were related to nudge acceptability. Because the majority of nudges were accepted, I opted to investigate which factors were related to people opposing the nudges. Accordingly, I dichotomised the 17 dependent variables, with one category denoting opposition, combining the ‘strongly disagree’ and ‘disagree’ responses, and the other denoting no-opposition, combining the ‘strongly agree’, ‘agree’ and ‘neutral’ responses.

Table 9 shows the coefficients of the logistic regression models, including all the relevant predictors of acceptability. The model for each nudge includes the subjective perception of effectiveness and manipulation, sociodemographic variables (age, gender, education and religious affiliation), political variables (ideology, trust in institutions, opinion on government intrusion with individual choices to improve individual wellbeing and government intrusion on individual choices to improve social wellbeing) and two of the behavioural variables (the subjective perception of situationism vs dispositionism for personal and other choices). In addition, the models include the behavioural variables related to each nudge. For instance, the models on the acceptability of default organ donation and the organ donation campaign include information on whether subjects are organ donors. The section 2 of Appendix A includes information on data preparation and the variables and categories included in the models. Multicollinearity was controlled using the variance inflation factor.

**Table 9. Logistic regression models predicting no-acceptability of the 17 nudges.**

	Default organ donation	Organ donation campaign	Default energy use	Social comparison of energy use	Default donations to charity	Charitable donations campaign	Carbon reduction app	Optical illusions to control speeding	Campaign to control speeding
(Intercept)	-2.884(1.88)	-3.205(2.66)	-1.191(1.50)	-1.880(1.28)	-1.172(0.93)	0.585(1.18)	0.432(1.38)	-0.859(1.36)	0.821(1.90)
Effectiveness (n)	3.489(1.20)**	3.284(0.85)***	4.188(0.56)***	3.938(0.48)***	2.123(0.31)***	2.861(0.39)***	3.556(0.58)***	3.457(0.44)***	4.412(0.83)***
Effectiveness (effective)	2.077(0.58)***	5.675(0.86)***	6.736(0.74)***	5.990(0.80)***	3.206(0.33)***	4.338(0.66)***	5.048(1.08)***	6.131(1.06)***	5.791(0.90)***
Manipulation (n)	0.283(0.81)	1.006(0.94)	1.721(0.60)**	1.143(0.52)*	0.0806(0.41)	-0.058(0.47)	-0.687(0.49)	0.284(0.45)	1.538(0.71)*
Manipulation (effective)	0.142(0.60)	-1.093(0.73)	-1.448(0.52)**	-1.010(0.41)*	-1.502(0.32)***	-1.626(0.39)***	-1.244(0.49)**	-0.962(0.42)*	-0.920(0.56)
Age	0.028(0.01)	0.040(0.02)	-0.021(0.01)	0.001(0.01)	0.004(0.00)	-0.014(0.01)	-0.012(0.01)	0.006(0.01)	-0.024(0.01)
Gender (F)	1.152(0.60).	0.148(0.76)	-0.672(0.47)	-0.589(0.38)	-0.242(0.26)	-0.134(0.36)	-0.258(0.44)	-0.419(0.41)	0.450(0.54)
Secondary education	0.875(0.60)	1.389(1.13)	-0.869(0.47)	1.175(0.65).	-0.088(0.45)	0.527(0.54)	1.603(0.64)*	0.501(0.61)	0.027(0.83)
University education	0.765(0.85)	2.002(1.13).	-0.703(0.69)	1.520(0.64)*	-0.036(0.44)	0.530(0.53)	2.414(0.64)***	0.508(0.60)	-0.318(0.83)
Religion (atheist/agnostic)	0.736(0.59)	0.831(0.76)	-0.218(0.47)	-0.027(0.39)	-0.145(0.27)	-0.321(0.34)	0.422(0.44)	-0.693(0.39)a	-0.160(0.50)
Trust in institutions (n)	1.026(0.64)	0.831(0.84)	0.472(0.45)	0.159(0.41)	0.401(0.27)	0.114(0.36)	0.363(0.44)	-0.395(0.40)	0.702(0.58)
Trust in institutions (h)	0.571(0.83)	-1.579(1.01)	0.079(0.66)	-0.669(0.54)	0.464(0.37)	0.426(0.57)	-0.307(0.64)	-0.191(0.59)	-0.926(0.76)
Ideology (Centre)	-2.135(0.77)	-1.861(1.04).	0.847(0.51)	-0.577(0.40)	0.080(0.28)	-0.455(0.38)	-0.750(0.50)	0.105(0.41)	-2.322(0.66)***
Ideology (Right)	-0.307(0.94)	-1.141(1.30)	-0.205(0.64)	-1.379(0.58)*	0.063(0.39)	0.091(0.52)	-1.674(0.63)**	-0.664(0.54)	0.535(0.81)
Interference pro-self (n)	-0.709(0.75)	-1.530(0.90).	-0.889(0.63)	0.213(0.54)	0.560(0.37)	0.059(0.49)	1.130(0.55)*	-0.296(0.50)	0.231(0.65)
Interference pro-self (s)	-0.707(0.86)	-1.909(1.13).	-0.740(0.66)	0.609(0.59)	0.540(0.39)	-0.224(0.55)	0.460(0.62)	-0.644(0.56)	-0.431(0.79)
Interference pro-social (n)	1.280(0.71)a	3.004(1.02)**	0.706(0.65)	0.486(0.58)	0.342(0.40)	0.651(0.52)	0.302(0.57)	1.257(0.54)*	1.882(0.70)**
Interference pro-social (s)	2.640(0.82)**	4.097(1.18)***	0.582(0.67)	0.308(0.60)	0.475(0.40)	1.137(0.56)*	0.471(0.63)	1.160(0.56)*	3.271(0.92)***
Situationism - self (n)	-0.431(0.56)	-0.836(0.72)	-0.789(0.47).	-0.190(0.41)	0.392(0.27)	-0.211(0.37)	-0.476(0.42)	-0.126(0.39)	-0.755(0.56)
Situationism self (h)	0.452(1.02)	1.881(1.38)	0.456(0.64)	-0.629(0.54)	0.277(0.38)	-0.376(0.48)	0.369(0.63)	-0.014(0.57)	0.523(0.87)
Situationism-others (n)	2.933(0.84)***	2.505(1.01)*	2.179(0.66)**	-0.594(0.55)	-0.590(0.42)	0.146(0.51)	-1.106(0.68)	-0.059(0.58)	-0.119(0.78)
Situationism-other (h)	1.752(0.69)*	1.164(0.83)	2.214(0.64)***	0.277(0.51)	-0.428(0.40)	-0.071 (0.47)	-0.490(0.65)	-0.008(0.52)	-0.255(0.77)
Organ donors	0.045(0.66)	-0.939(0.90)							
No concern environment			-0.998(0.48)*	-0.042(0.43)			-0.872(0.40)*		
Not used to donating					-0.157(0.27)	-0.451(0.34)			
Non-polluting transportation							0.203(0.39)	0.476(0.35)	0.855(0.48).
No traffic fines								-0.113(0.41)	-0.704(0.63)

Logistic regression models. The non-approve category taken as reference (0). Probability of 1 over 0. Standard errors in parentheses. All categorical predictors in exception of age included as a continuous variable. Different behavioural factors included for each Nudge. (n) = neutral, (s) = support, (h)=high

Significance. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

(Table 9 continues)	Tax compliance campaign	Default food layouts	Traffic light labels	Automatic enrolment for social benefits	Gambling control app	Gambling control campaign	Regulation of alcohol container sizes	Voting reminders
(Intercept)	0.274(1.07)	-0.922(1.50)	-0.562(1.68)	0.451(1.52)	1.664(1.54)	-0.849(1.57)	-1.467(0.92)	-1.172(1.08)
Effectiveness (3)	2.125(0.57)***	2.525(0.48)***	3.337(0.62)***	2.381(0.52)***	2.742(0.54)***	2.415(0.55)***	3.039(0.37)***	3.695(0.40)***
Effectiveness (4-5)	17.462(689.53)	4.712(0.82)***	4.316(0.65)***	5.508(0.68)***	4.289(0.79)***	4.262(1.08)***	4.869(0.77)***	6.415(1.06)***
Manipulation (3)	0.459(0.43)	1.422(0.62)*	0.188(0.61)	0.010(0.54)	0.919(0.60)	1.185(0.61)	0.464(0.32)	-0.281(0.38)
Manipulation (4-5)	-0.742(0.40)	-0.855(0.49)a	-0.373(0.61)	-1.285(0.49)**	-1.139(0.46)*	0.339(0.57)	-0.948(0.31)**	-1.426(0.39)***
Age	0.006(0.01)	0.013(0.01)	0.02(0.01)	0.026(0.014)	-0.0170(0.01)	0.019(0.015)	0.002(0.00)	-0.00(0.01)
Gender (F)	-0.476(0.32)	1.779(0.48)***	0.609(0.48)	-0.280(0.42)	0.353(0.46)	0.063(0.47)	0.225(0.27)	-0.104(0.33)
Secondary education	0.494(0.52)	0.065(0.68)	0.477(0.75)	-1.289(0.77)	0.050(0.74)	-0.162(0.68)	0.288(0.46)	0.248(0.56)
University education	0.774(0.52)	0.427(0.68)	-0.256(0.74)	-1.945(0.77)*	-0.325(0.73)	1.091(0.74)	0.287(0.45)	0.823(0.54)
Religion (atheist/agnostic)	-0.165(0.32)	0.081(0.44)	0.468(0.475)	-0.106(0.42)	0.420(0.43)	0.097(0.44)	-0.323(0.27)	0.055(0.33)
Trust in institutions (n)	0.821(0.35)*	-0.251(0.46)	0.736(0.54)	0.077(0.46)	-0.177(0.46)	-0.344(0.47)	0.164(0.29)	0.748(0.35)*
Trust in institutions (h)	0.585(0.48)	-0.370(0.63)	-0.440(0.66)	-0.356(0.63)	-0.155(0.61)	-0.2522(0.69)	0.272(0.36)	-0.660(0.44)
Ideology (Centre)	-0.370(0.35)	-0.905(0.50)	-1.092(0.57)	-0.1992(0.49)	-0.070(0.46)	-1.107(0.55)*	-0.061(0.29)	-0.624(0.35)
Ideology (Right)	0.165(0.51)	-0.348(0.62)	0.07(0.70)	-1.351(0.58)*	0.648(0.68)	-0.793(0.71)	0.142(0.40)	-0.00015(0.48)
Interference pro-self (n)	-1.013(0.44)*	0.697(0.55)	0.142(0.61)	0.0143(0.59)	-0.423(0.53)	0.064(0.57)	0.225(0.38)	0.888(0.48)
Interference pro-self (s)	-0.167(0.52)	0.286(0.64)	-0.975(0.72)	-0.065(0.64)	0.503(0.63)	0.048(0.67)	-0.015(0.40)	0.992(0.49)*
Interference pro-social (n)	1.163(0.52)*	1.310(0.62)*	1.080(0.64)	-0.778(0.65)	0.093(0.59)	1.595(0.65)*	0.588(0.44)	-0.094(0.52)
Interference pro-social (s)	0.744(0.47)	1.344(0.63)*	1.78(0.72)*	0.246(0.68)	1.133(0.62)	0.725(0.65)	0.515(0.44)	0.870(0.52)
Situationism - self (n)	0.015(0.33)	0.104(0.44)	-0.06(0.49)	-0.047(0.47)	-0.122(0.44)	0.234(0.48)	-0.577(0.28)*	0.391(0.344)
Situationism self (h)	0.299(0.50)	0.258(0.67)	-0.617(0.61)	0.472(0.63)	0.039(0.71)	1.071(0.79)	-0.716(0.42)	0.212(0.47)
Situationism-others (n)	-0.251(0.49)	-0.734(0.67)	-1.339(0.77)	0.891(0.67)	-1.322(0.74)	-0.177(0.66)	0.325(0.44)	-1.038(0.56)
Situationism-other (h)	0.093(0.47)	-0.405(0.65)	-0.351(0.78)	0.482(0.61)	-1.3877(0.71)	0.434(0.61)	0.587(0.42)	-1.159(0.55)*
Does not do tax forms	0.023(0.50)							
Tax form done by third party	0.075(0.46)							
Average healthy diet		-0.170(0.41)	-0.486(0.47)					
Unhealthy diet		0.365(0.88)	0.249(0.99)					
No help to eat healthily		-1.007(0.42)*	-0.316(0.44)					
No social benefits				-0.181(0.65)				
No gambling					0.640(0.56)	-0.071(0.64)		
Occasional drinker							-0.535(0.37)	
Regular drinker							-1.065(0.32)**	
Heavy drinker							-0.556(0.43)	
Does not always vote								-0.487(0.49)

The logistic regression models revealed that several independent factors were related to nudge acceptability. Firstly, across all interventions, the strongest predictor was the perceived effectiveness of the intervention. As expected, the relationship was positive. For all nudges, people that believed that the intervention would be effective were less likely to oppose the nudge. The result of the study confirmed hypothesis 5 and coincided with previous research on the acceptability of nudge interventions, in which perceived effectiveness was found to be a relevant factor for accepting nudges (Arad & Rubinstein, 2018; Davidai & Shafir, 2018; Petrescu, et al. 2016; Sunstein, 2017c).

Following perceived effectiveness, the perception of manipulation of nudges was also relevant. Perceived manipulation affected the acceptability for 12 out of the 17 nudges. As expected, for almost all cases, the effect was negative. Respondents that believed that a nudge was manipulative were more likely to disapprove of its use. The effect was relevant for automatic nudges; it affected the acceptability of the default to cut energy use in hotels, the use of defaults on tax forms to encourage donations to charity, the automatic enrolment of citizens eligible for social benefits, the app to control gambling and the nudge altering the size of alcohol containers. The effect, however, was also noted for some System 2 interventions, specifically for the control of energy use by social comparison, the tax compliance campaign, the campaign to promote donations to charity, the carbon reduction app, the campaign to control speeding, and the use of voting reminders. Overall, the results confirm hypothesis 6; the effect of perceived manipulation was relevant and negatively affected support for nudges.

The effect of sociodemographic factors was minimal and only relevant for a few nudges. Gender affected support for default food layouts in restaurants and supermarkets to promote healthy eating, with men more likely to disapprove of the intervention than women. The higher opposition among men to using default food layouts was probably related to the fact that men tend to be less concerned about body size and weight. The result was consistent with previous studies that found higher support for nudges related to eating behaviour among women (Evers et al., 2018). Education affected the support for three interventions but in different directions. Education affected support for two pro-environment interventions: people with a university education were less likely to oppose the use of social comparison to reduce energy consumption in households, and people with a secondary and university education were less likely to oppose the use of an add-in in Google maps to reduce carbon emissions. The result concurs with previous studies that have documented the effect of educational attainment on pro-environment policy interventions (Clery & Rhead, 2013; Meyer, 2015). By contrast, people with a university education were more likely to disapprove of the automatic enrolment of citizens for social benefits. A possible explanation for this result might be that they perceived themselves to be less likely to need or access social benefits. Overall, the effect of sociodemographic variables was minimal and appeared to be related to the aim of the specific nudges. The results confirmed hypothesis 7 and indicated that sociodemographic factors did not systematically affect preferences for nudges.

A mixed picture emerged from considering the effect of political factors. Firstly, as expected, ideology did not have a systematic effect on nudge acceptability. People who considered themselves to be politically centrist were more likely to oppose the campaign to reduce speeding and the campaign to control gambling. Right-wing people were more likely to disapprove of two pro-environment interventions: the use of social comparison to reduce energy consumption at home and the app to reduce transport emissions. They were also more likely to oppose the automatic enrolment of citizens for social benefits. Again, the effect of ideology seemed to be primarily motivated by the aim of specific nudges. Conservatives (right-of-centre) were against intervention with politically contested aims such as environmental protection and social benefit programmes. Interestingly, people who considered themselves to be politically centrist were more likely to oppose campaigns with negative emotional messages. Overall, the effect of ideology was not systematic but related to the policy goals, results which are consistent with hypothesis 8.

Contrary to previous studies, trust in public institutions did not have a systematic effect on the acceptability of nudges. It did, however, affect support for two interventions in which trust in public institutions appears to be especially important. Those that distrust institutions were more likely to disapprove of the use of social and moral norms to improve tax compliance, and the use of reminders to increase voter turnout. Similarly, preferences for the government interfering in individual choices for promoting wellbeing had a very modest effect and only affected support for three nudge interventions. People who supported government interference in individual choices for improving individual wellbeing were less likely to reject the app to reduce transport emissions, the tax compliance campaign and the voting reminders. I expected support for interference in promoting individual wellbeing to have a more relevant effect on the support for pro-self nudges; however, this was not the case. Only one of these three interventions involved pro-self nudges; yet tax compliance and environmental protection also have positive implications at an individual level, which may explain the effect of the variable.

After perceived effectiveness and perceived manipulation, support for government interference in individual choices for improving social wellbeing was the more reliable predictor of non-opposition to nudges. As expected, the variable had a positive effect, and those in favour of government interference for social good were less likely to reject nudges. The effect was significant for 10 out of 17 nudges. In the case of pro-social nudges, people who supported interference for social good were less likely to reject default organ donation, the organ donation campaign, the campaign to promote charitable donations and the tax compliance campaign. In the case of pro-self nudges, people in favour of interference were less likely to reject optical illusions on roads, the speeding control campaign, the default food layouts, the campaign to control gambling and the voting reminders.

The results on the effects of trust in public institutions and support for government interference in individual choices both confirmed and revoked hypothesis 9. On the one hand, no systematic effect was found on trust in institutions and support for interference in individual wellbeing for individual

good, findings that revoked hypothesis 9. The non-effect of these factors was not anticipated but could be explained by the fact that the models analyse opposition to nudges, rather than support for nudges, as was done in the previous studies that found a significant effect on trust in institutions. What is relevant is that, surprisingly, distrust in public institutions does not generally undermine the acceptability of nudges.

On the other hand, support for government interference in individual choices for the social good had an effect on several nudges. The result was interesting because it indicated that people's views about government interference in individual choices in different areas and decisions were relevant in terms of forming opinions on nudges. The finding is consistent with previous studies that found greater support for nudges among people with communitarian views (Hagman et al., 2015) and a high level of opposition to nudges among individualists (Jung & Mellers, 2016). In essence, these findings support hypothesis 9 and suggest that more empirical research is needed to clarify in which areas and under what conditions people support governments interfering in their choices to improve social wellbeing.

The models also analysed the effect of behavioural factors. Firstly, they included a perception of dispositionism vs situationism with regard to personal behaviour and the behaviour of others. These two factors did not have relevant effects. The perception of self-situationism was only related to the nudge on limitations on the size of alcohol containers, with those that perceived themselves as being more sensitive to the influence of the environment being more likely to reject this nudge. This result may be explained by the fact that people with a perception of self-situationism might perceive this intervention as being more likely to affect their behaviour against their will; thus, they rejected it more. The perception of others' situationism was a bit more relevant and affected the acceptability of four interventions. The data revealed that people who believed that the behaviour of others is strongly influenced by the environment were less likely to reject default organ donation, the organ donation campaign and the app to reduce energy use in hotels. By contrast, they were more likely to reject the use of voting reminders. A possible explanation for this difference is that people who perceived other's situationism might be more likely to accept nudges when they accept the aim of the nudge, but less likely to accept nudges when they promote an aim that they perceived goes against the interest of most choosers. Indeed, voting reminders was one of the three nudges that was rejected. Overall, the findings on personal and others' perception of situationism vs disposition were somewhat limited in terms of confirming hypothesis 11. While the variables were significant for some nudges, their effect was not general. The effect of these factors was an issue that further research needs to explore in greater depth.

Finally, the survey collected information on personal habits and behaviour to study how they affected nudge acceptability. As explained, the models included relevant behavioural factors for each nudge. In line with previous studies, people's personal preferences and habits had an impact on the acceptability of some nudges, in both a positive and negative way. Firstly, people who expressed no concerns for the environment were more likely to disapprove of the default to reduce energy consumption in hotels and

the app to reduce carbon emissions. Likewise, people who indicated that they regularly drank alcohol during the week were more likely to disapprove of the nudge limiting the size of alcohol containers. The findings were in line with previous research that found that those groups that were targeted by the nudge were less likely to approve them (Evers et al., 2018). The results also indicated that preferences can have a positive effect. In fact, people who expressed problems maintaining a healthy diet were less likely to oppose the default food layouts in restaurants and supermarkets. The results suggested a tendency to reject nudges that did not align with personal preferences, thus confirming hypothesis 12. However, the effect was not observed for all the nudges, and the findings were insufficient to confirm hypothesis 12. Further research needs to carry out an in-depth exploration of this hypothesis, considering how personal preferences affect the acceptability of nudges in various areas of intervention.

## **5. Discussion**

The study found that support for nudges in Spain is high. The majority of interventions were supported and had moderate to high levels of support. People approved of the use of different types of nudges and accepted their use in several policy areas. However, three of the seventeen nudges were disapproved of by the majority. The results resemble the findings in other European countries and western democracies and suggest that Spain belongs in the “principled pro-nudge nations” category, in which studies document general support for nudge interventions, but opposition to specific nudges (Sunstein et al., 2018).

The main factor affecting the acceptability of nudges was their policy aim. The variability in support for the nudges presented was related to the aim they promoted, and the rejection of interventions appeared to be driven by the fact that people disagreed with their purpose. The findings concurred with previous studies that indicated that support for nudges varied significantly depending on their policy goals (e.g. Jung & Mellers, 2016; Sunstein, 2017b).

The survey included two questions on people’s perception of two specific issues regarding nudges: whether they were seen as effective and whether they were seen as manipulative. In general, respondents had an unclear perception of both issues, perhaps because they were unfamiliar with the interventions or because the survey did not include any information about these two aspects. However, nudges were not seen as intrinsically ineffective or intrinsically manipulative, and people were closer to finding them effective and non-manipulative than anything else.

The findings indicated that people had positive attitudes towards the implementation of nudges, accepted their use and rejected them when they disagreed with the aim rather than based on any intrinsic characteristic or general impression about their potential ineffectiveness and potentially manipulative nature. Overall, the findings showed that people valued nudges as they would value any other policy instrument. The results were favourable to the implementation of nudges in Spain and suggest that

nudges will be broadly accepted. However, the survey also documented interesting issues that are specific to nudges and that have relevant implications for their use in policymaking.

The type of policy goal (pro-self vs pro-social) and the type of nudge (automatic vs reflective) are crucial factors in assessing nudge acceptability.

Examining the effect of the type of policy aim, the results were in line with previous studies and indicated that people prefer pro-self nudges over pro-social nudges. The findings agreed with the original nudge characterisation and the idea that nudges should be used to correct externalities and improve an individual's wellbeing. However, in Chapter 2 and 4, I discuss the problems of justifying nudges extensively under the argument that they improve people's wellbeing and argue that they have more potential and less ethical problems when they are used to promote social good. The results suggested that people liked nudges that improve individual wellbeing and contradicted my argument. However, the results also indicated that favouritism for pro-self nudges should be taken with caution.

Pro-social nudges are applied in areas of public policy that have a social dilemma structure, and in which common good requires some type of sacrifice or cost for individual wellbeing. Likewise, pro-social nudges target everybody. By contrast, pro-self nudges target groups that engage in specific behaviours and do not pose any barrier to those not targeted by the nudge. Accordingly, I speculated that favouritism for pro-self nudges does not reflect people's acceptance of nudges that affect themselves, but of nudges that improve the behaviour of others.

Some of the survey's results supported these ideas. For instance, occasional drinkers were more likely to reject the nudge to reduce alcohol consumption; support for nudges to control gambling was high, yet only 2% of the sample were regular gamblers; support for nudges to promote healthy eating was also higher; yet it was lower among people who did not perceive themselves as being in need of help to maintain a healthy diet. Likewise, while pro-social nudges had more moderate levels of support on average than pro-self nudges, two pro-self nudges were disapproved of by the majority. In this sense, the data showed that people disapproved of pro-self nudges if they felt that they promoted illegitimate ends. Finally, support for interference in individual choices for the social good was a relevant predictor of acceptability for both pro-self and pro-social nudges. An implication of this is the possibility that people perceived both pro-self and pro-social nudges as matters of public good rather than individual welfare.

People might have favoured pro-self nudges not because they accepted their use in changing their behaviour but because they were not bothered by them. These ideas are in line with previous studies that found greater support for pro-self nudges among people not targeted by the nudge and greater support for all types of nudges among women and empathetic people, possibly due to the fact that they worry more about the behaviour of others (Evers et al., 2018; Gyrd-Hansen & Kjær, 2015; Jung & Mellers, 2016).

The finding, while merely preliminary, suggests that support for pro-self and pro-social nudges follow different standards. The acceptability of pro-social nudges might be related to how much people care about a particular policy goal. For instance, the study found that people who were concerned with environmental protection were less likely to oppose nudges for environmental protection. Likewise, support for pro-social nudges might be higher in cases where they target policy goals that are already broadly accepted, such as tax compliance. Interestingly, I found that support for the use of a default to promote organ donation was higher than in other countries, perhaps because Spain is the world leader in organ donations and transplants, and people are aware of the importance and success of the policies to promote organ donation. By contrast, to assess the acceptability of pro-self nudges, it might be more relevant to understand in which cases people perceived that nudges targeted their behaviour and how those targeted by the nudge felt about their implementation. As the data shows, people's personal preferences and habits had an impact on the acceptability of some nudges, in both positive and negative ways; this is an issue that requires further empirical exploration.

When examining which types of nudges people preferred, I found that public acceptability of nudges was conditional on the nature of the nudge. Support for reflective nudges was higher than support for automatic nudges. Likewise, people perceived automatic nudges as being more effective and more manipulative, which indicated that they had an intuitive and somewhat correct idea of the different level of intrusiveness of nudges. However, opposition to automatic nudges was not categorical. In agreement with the ideas presented in Chapters 2 and 5, the data suggests that support for automatic nudges depends on the context in which they were applied and the aim they promote.

Regarding the acceptability of automatic nudges, the findings point to three different aspects. Firstly, people appear to approve of automatic nudges when they are used to promote ends that they perceive as legitimate, such as in the case of healthy eating; or when they are used to improve the design and performance of policy programmes, as shown in the support for the automatic registration of citizens eligible for social benefit programmes. Secondly, the results also suggest that the acceptance of automatic nudges is not only affected by whether people agreed with their aim but by the context in which they are used. For instance, respondents seemed to favour the use of automatic nudges in contexts in which people might have self-control problems, such as gambling. By contrast, people disliked automatic nudges in contexts with moral connotations, such as organ donation, or in contexts where they like to retain control, for instance, driving behaviour. In both cases, while people supported the aims of the nudges, they were less in favour of employing automatic nudges. Finally, in accordance with previous studies, people dislike the use of automatic nudges in matters that involve some kind of loss, particularly economic losses, as shown by the opposition to the default setting of donations to charity and the more significant opposition to the limitation on the size of alcohol containers (Jung & Mellers, 2016; Sunstein, 2017b). In essence, support for automatic nudges is trickier. Further work is

required to explore and understand the interaction between policy goals, nudge types and context of use, which appears to be crucial to assessing the acceptability of automatic nudges.

Overall people reacted differently to nudges depending on the type of policy goal and the type of nudge. In this sense, revising attitudes towards nudges as a whole has value. However, because nudges are so diverse, both in form and in aim, future research should separately consider how people feel about pro-self nudges and pro-social nudges, and also how they feel about automatic and reflective nudges.

With regard to the independent predictors of acceptability, the results had important implications for the implementation of nudges in policymaking. Sociodemographic factors and political factors such as ideology and trust in institutions had minimal effects, and the results suggested that their effects appear to be related to the purpose of specific nudges rather than with the nudges themselves. The findings indicated that all types of citizens will accept the use of nudges and suggested that no particular groups will reject their implementation.

The findings where perceived effectiveness and perceived manipulation influenced nudge acceptability also had important implications for the use and communication of nudges. The observed association between perceived effectiveness and perceived manipulation and nudge acceptability reveals that people favoured interventions that they perceived as being effective and opposed nudges that perceived as being manipulative. However, the results also revealed that people had unclear beliefs about both factors and in some cases, misjudged them. For instance, the nudge limiting the size of alcohol containers was perceived as the most ineffective intervention, yet previous research indicated that limiting portion sizes tends to be effective in reducing consumption (Hollands et al., 2015; Kersbergen et al., 2018). Likewise, my expectations on people's perceived manipulation did not wholly match people's actual perception. For instance, the use of social comparison to reduce energy consumption and the use of voting reminders were perceived as being more manipulative than I anticipated.

In the absence of real information on both issues, people's perceptions will affect how they value and accept a nudge. Accordingly, research should explore how the communication or framing of the effectiveness and nature of nudges affects their public acceptability. Previous studies documented that information about effectiveness helps to improve the acceptability of nudges, especially automatic nudges (Davidai & Shafir, 2018; Gyrd-Hansen & Kjær, 2015; Petrescu, et al., 2016; Sunstein, 2017c). Likewise, some studies found that using different frames to describe the influence of nudges (for instance, as conscious or unconscious) affects the acceptability of several interventions (Felsen et al., 2013; Jung & Mellers, 2016; Sunstein, 2017c). Findings on both issues, however, are still limited. Future research should explore this association further and look at how to implement and communicate nudges to ensure their acceptability.

Another important finding was that support for government interference in people's choices for social good affected the acceptability of 10 out of 17 nudges. The result supports the idea that the domain in

which nudges are applied is relevant to their acceptability. The notion that individual choices generate unsatisfactory social outcomes might be more apparent in some areas, for instance, tax compliance, but might be unclear in others, such as occasional alcohol consumption. In this sense, the acceptability of nudges will be high in those contexts in which people already recognise the importance of limiting people choices, and in those cases in which nudges are used to support policy goals that are already relevant and accepted.

The results also suggest that nudges may have the potential to introduce intervention in areas in which people reject interference in individual choices. Previous research suggests that gradual interventionism, from softer to more coercive measures, is useful in regulating behaviour in areas in which people oppose any kind of interference (Bicchieri, 2016). In addition, some studies suggest that people favour the use of nudges over more coercive paternalistic interventions. In this sense, nudges can be used to introduce regulation in areas with traditionally limited interventionism. Finally, the association between support for interference in individual choices for social good and nudge acceptability also provides interesting insights into how to communicate nudges. In some areas, people might be unaware of how individual behaviour might contribute to social problems. In these cases, framing interventions in ways that make salient the relation between personal behaviour and social wellbeing more salient may be helpful to increasing both the perception of responsibility and the acceptability of nudges.

The perception of others' situationism was relevant to predicting the acceptability of four interventions. The data revealed that people who believed that others' behaviour was highly influenced by their environment were less likely to oppose 4 out of 17 nudges and more likely to oppose one of the less liked nudges. It is possible that the perception of other's situationism makes people support some nudges because they are aware of the need to interfere in people's choices to change behaviour. However, this same belief also makes them more likely to oppose nudges that they perceived as going against the wellbeing of others.

As argued in Chapter 2, I believe the context of choice and, specifically, the nature of the decision in different areas to be a crucial element in analysing how nudges will change behaviour and assess their ethical implications. Depending on the context, people's choices might be more dispositional or, instead, more driven by the situation and, in such cases, nudging might have more positive than negative implications. The results moderately indicate that acknowledging the influence that the environment has in the behaviour of others affects the acceptability of nudges both in positive and negative ways. Previous studies found similar results. For example, Petrescu et al. (2016) found that the belief that the environment is responsible for people being overweight increases the acceptability of interventions to reduce the consumption of sugar-sweetened beverages. However, Gyrd-Hansen & Kjær (2015) found that acknowledging problems of self-control in health-related behaviours did not increase support for government interference in those areas.

Interestingly, the results show that people might hold inaccurate beliefs about how choice architecture influences their choices. A large percentage of respondents (about 50%) stated that their decisions were made following their preferences and dispositions rather than being influenced by their environment. However, many (about 47%) believed the behaviour of others to be mostly driven by the context of choice. This discrepancy might suggest that people understand the relevance that their environment has in shaping behaviour; however, they do not acknowledge how it impacts on themselves. This finding, while preliminary, suggests that there is room to inform and influence people's subjective perception of how different environments influence their choices and that this factor will shape their acceptability of different nudges. Further work is required to establish the association between the nature of the decision in different environments, the subjective perceptions of how people make choices and the acceptability of interference in these different domains.

Finally, it is worth noting some of the study's limitations. Firstly, the survey collected attitudes towards hypothetical interventions. The information is interesting when used to consider how people feel about nudges and helps to generally guide the implementation of specific nudges in different policy areas; however, the results are insufficient to understand how people will react to their real-world implementation. Secondly, the survey is useful to capture a general impression of public acceptability of nudges and at identifying general themes about which nudges people prefer and where and how to use them. However, the survey was unable to provide answers to some relevant questions. For instance, while the survey revealed a general preference for pro-self and reflective nudges, the data did not demonstrate which factors cause these differences. Similarly, while the survey showed the association between different predictions and nudge acceptability, using the available data it is not possible to fully grasp why and in which ways these factors affected public support for nudges. Many of the unresolved questions require a more specific look and the use of experimental methods to identify the causal relationship between factors.

## **6. Concluding remarks**

This chapter has analysed attitudes towards nudges in Spain using a nationally representative survey that collected opinions on different nudges applied to diverse policy aims. The chapter contributes to the general discussion on public attitudes towards nudges and introduces the perspective of a new context in which nudges are still underused. The present study documents general support for nudge interventions in Spain and provides empirical insight into how different factors influence people's acceptability of diverse interventions. The results were in line with previous studies on attitudes towards nudges, document issues to consider for their implementation in policymaking and serve to identify factors that need to be explored further by future research. Overall, the findings suggest the implementation of nudges in Spain will be accepted and provide a preliminary outline of how different elements of their implementation should be.

## Chapter 7

# Preferences for automatic and reflective nudges: experimental evidence and implications

### 1. Introduction

Much of the discussion on the acceptability of nudges is associated with how nudges affect people's behaviour. As extensively discussed in Chapter 2, several objections to the use of nudges come from the idea that nudges affect behaviour in ways that undermine people's autonomy. Critics argue that nudges change behaviour by bypassing rationality or reflection and, therefore, they compromise people's autonomy. An important response to this criticism has been the clarification that nudges differ in the way that they affect people's choices. While different classifications exist, the consensus is that some nudges operate in more transparent and reflective ways (reflective nudges) while other nudges are less noticeable and less engaging with conscious decision-making processes (automatic nudges). Using this differentiation, the tendency has been to argue that automatic nudges are morally problematic while reflective nudges are mostly unproblematic.

Along similar lines, literature on public acceptability of nudges has explored how people feel about different types of nudges. As discussed in Chapter 5, the findings indicate that people favour interventions that are more transparent and reflective over nudges that work covertly and limit deliberation. The findings of the different studies about this issue indicate that negative attitudes towards automatic nudges are motivated by concerns about intrusiveness and manipulation, which suggests that people also share the concerns expressed in the philosophical discussion. Consistent with these findings and as described in Chapter 6, I found that in Spain, automatic nudges are less accepted and perceived as more manipulative than reflective nudges.

Research on public attitudes towards nudges shows general preferences for reflective nudges. However, people do not categorically oppose the implementation of automatic nudges. In fact, they actually support their use in several cases. Different studies indicate that the acceptability of automatic nudges depends on their goal, the context in which they are applied, and how they are communicated to subjects. Therefore, it is necessary to explore how these factors influence positive and negative attitudes towards these interventions.

Chapter 7 focuses on exploring preferences for automatic nudges. The chapter investigates how the acceptability of automatic nudges varies depending on the context in which the nudge is applied and how the nudge is communicated. To that end, two online experiments have been conducted: Study A is a survey experiment that explores whether preferences for automatic and reflective nudges vary

depending on the information provided about the nudges and the time available to subjects to evaluate the interventions. Study B is a contrastive vignette experiment that explores the acceptability of automatic and reflective nudges depending on the type of decision that the nudge is intended to change and the domain in which the nudge is applied. The studies show that the acceptability of automatic nudges depends largely on the aim they are promoting, and people support their implementation if they approve of their aim.

Chapter 7 proceeds as follows. Section 2 presents the research questions and hypothesis that guide the empirical analysis. Section 3 details the methods and the results of Study A, and Section 4 details the methods and the results of Study B. Section 5 is dedicated to a general discussion and explains the main themes regarding the acceptability of automatic nudges. The chapter ends with the conclusions.

## **2. Research questions and hypothesis**

Research on public support for nudges indicates that people prefer reflective nudges to automatic nudges; however, opposition to automatic nudges is not clear-cut. In some cases, people judge automatic nudges to be problematic due to how they work; however, people do not oppose their use categorically. Firstly, elements of the design and implementation of nudges affect their overall acceptability. Consistent with the ideas discussed in Chapter 2, studies show that attitudes towards automatic nudges are affected by the aim and the context in which they are applied. For instance, research suggests that their acceptability increases when automatic nudges are used for good to ease decisions in situations of cognitive load, when they affect unconscious choices or when they do not affect choices in relevant domains. Secondly, attitudes towards automatic nudges are affected by how the nudges are communicated. Some studies indicate that the method of evaluation of nudges and the information available to subjects about them also influence whether they accept or reject their implementation.

Judging the acceptability of nudges involves considering whether and when it is acceptable to affect people's choices by triggering non-deliberative decision-making processes. It is a morally ambiguous situation that involves different and conflicting ideas and values of what is right or wrong. Accordingly, whether people would accept or reject their use requires exploring two different issues. Firstly, it requires analysing the situations in which, if any, people perceive that influencing choices in such matters is acceptable. Secondly, it requires exploring how the presentation of the nudges influences people's opinion about their implementation. The present studies investigate these issues with the following research questions and hypothesis.

- *Research question 1. Are preferences for automatic nudges affected by how nudges are described?*

Recent studies on the public support for nudges indicate that the way in which nudges are presented to subjects influences their opinions on how acceptable they are. As explained in Chapter 5, automatic

nudges are more disapproved of when directly compared to reflective nudges and the use of informational and educational campaigns. However, their acceptability increases when they are evaluated on their own or compared to more coercive measures such as taxes and regulations. Likewise, when people have additional information about nudges, for instance, how effective they are, or whether they work through conscious or unconscious mechanisms, their perceptions and opinions about automatic nudges also change. In line with these findings, some authors suggest that people's evaluation of nudges, particularly automatic nudges, varies depending on the way in which they are presented to subjects and which of the nudges' specific attributes are salient (Arad & Rubinstein, 2018; Davidai & Shafir, 2018; Sunstein, 2017b).

The acceptability of automatic nudges is morally ambiguous. From a deontological position, some may argue that influencing people's behaviour through non-deliberative mechanisms is always wrong, regardless of the consequences, the available alternatives, or the context in which these interventions are applied. By contrast, from a consequentialist position, it is possible to argue that the use of automatic nudges is permissible and positive if they ensure positive outcomes, are more cost-effective than other interventions and are used in contexts in which people tend to have already been influenced by unconscious and uncontrolled factors (e.g. Fischer & Lotz, 2014; Schmidt & Engelen, 2020; Schubert, 2015; Sunstein, 2015a). The fact that people are more suspicious of automatic nudges but still support their use in several instances indicates that they are aware of this trade-off. In some situations, people seem to think that the ends justify the means, while in other cases, they do not.

Whether judging the acceptability of automatic nudges involves trade-offs between different moral values is relevant when it comes to understanding the reasons behind approving or rejecting their implementation. Traditional approaches to moral judgment view and analyse ethical decision-making as a rational and deliberative process. However, alternative approaches emphasise that moral decisions do not always reflect deep moral convictions but tend to be affected by subconscious, intuitive and automatic processes. Haidt (2001) argues that moral judgment is mostly automatic, driven by moral intuitions that trigger feelings of good-bad or like-dislike that are later rationalised, resulting in moral reasoning. Greene et al. (2001) propose a dual-process theory of moral reasoning in which both conscious and subconscious factors affect ethical decisions and activate differently according to the personal involvement in a situation. Personal moral dilemmas trigger affective and automatic responses and result in deontological judgments. By contrast, impersonal moral situations trigger rational responses and result in consequentialist judgments. While there is no fully developed theory of moral decision-making, these approaches suggest that moral judgment is less intentional and reflective than previously conceptualised.

Research on experimental ethics supports the idea that moral judgments can be unreflective and automatic, and affected by ethical triggers, framing effects and moral heuristics. Several studies have found that conscious or subconscious primes affect moral judgment and ethical decisions. Welsh &

Ordóñez (2014) found that subconscious ethical and unethical priming triggers moral standards and reduces unethical behaviour in several tasks. Ong et al. (2014) found that the presentation of expressions of disgust primes individuals with high disgust sensitivity to more utilitarian judgments. Capraro et al. (2019) found that when intuitive thinking is primed, it favours deontological judgments, whereas when reasoned thinking is primed, it favours utilitarian judgments. Research also suggests that ethical triggers can be very diverse. For instance, Parzuchowski & Wojciszke (2014) found that the non-emotional gesture of hand-over-heart promotes honest responses. Lee et al. (2018) found that social observation increases deontological judgment in moral dilemmas due to reputational concerns. Shu et al. (2012) found that signing a form at the beginning instead of the end makes the ethical component of a task salient and decreases dishonesty in self-reports. Finally, research has also shown that deontological and utilitarian judgments may be affected by the time and resources available to agents to decide. Typically, utilitarian judgments require greater cognitive effort. Therefore, under conditions of cognitive load, fatigue or time pressure, people's judgment tends to be more deontological (Greene et al., 2008; Lütge et al., 2014; Suter & Hertwig, 2011; Timmons & Byrne, 2019).

Another important contribution to understanding the mechanisms underlying moral decisions is the notion of moral heuristics. Heuristics are simple rules of thumb that work well in providing an answer to complex tasks quickly, but that can easily misfire. The literature on heuristics and biases shows that employing heuristics such as representativeness or availability leads people to make incorrect assessments of probability. Sunstein (2005) argues that a similar process occurs when addressing moral issues. In decisions concerning moral and political domains, people might also rely on simple rules of thumb that might work well in general but that may also fail in certain contexts. Sunstein clarifies that moral heuristics are not linked to specific moral or philosophical preferences but can reflect different convictions, such as deontological and consequentialist values<sup>18</sup>. Sunstein argues that when we face a morally ambiguous situation, System 1 employs heuristics to classify the situation as being either right or wrong. However, while heuristics are effective in providing quick responses, they tend to be inadequate in many cases. In several contexts, System 2 might offer a different moral judgment.

The evaluation of automatic nudges might be affected by these types of factors. As argued, judging their acceptability involves different and contradictory moral values. Likewise, for the majority it is a

---

<sup>18</sup> Sunstein (2005) provides a few examples of moral heuristics. For instance, as far as risk regulation is concerned, people tend to punish companies that base their decisions on cost-benefit analysis, even if they place high value on human life. By contrast, they impose less severe punishments on companies that are willing to assess risks on people, without considering risk. Sunstein argues that the heuristic of 'not knowingly causing human death' applies in this situation, whereby the action of the company doing the test feels more wrong than the action of the company not doing the test. A second example of moral heuristics is that people tend to punish harm resulting from action more than harm resulting from omission. While direct action is generally more harmful than omission, in some cases it is not. However, due to the fact that 'acts are worse than omission heuristic', we fail to recognise when the distinction is relevant. For instance, euthanasia is highly controversial, yet not providing additional treatments to terminal patients is general practice in several countries.

new topic with which they are not familiar. As a result, subjects might be more influenced by how nudges are presented and more prone to using moral heuristics to determine whether they approve of or reject their implementation. Davidai & Shafir (2018) found that attitudes towards automatic nudges are affected by the mode of evaluation. When people evaluate the acceptability of automatic nudges in direct comparison to reflective nudges, they are overly influenced by the deliberative – non-deliberative difference. As a result, opposition to automatic nudges increases. However, when subjects evaluate them separately, the acceptability of automatic nudges increases. Similarly, some studies indicate that concerns regarding manipulation and intrusiveness on freedom of choice or autonomy are more likely to arise when the non-deliberative nature of automatic nudges is highlighted, yet less frequent when it is not explicitly described (Felsen et al., 2013; Jung & Mellers, 2016). By contrast, attitudes towards automatic nudges are more positive when the description of them emphasises their effectiveness (Sunstein, 2017c). In these cases, the positive information about nudges primes a favourable judgment.

In the task of evaluating the acceptability of automatic nudges, people might be using simple moral heuristics such as: ‘affecting behaviour through non-deliberative mechanisms is manipulative’ or ‘do not influence people’s choices without their awareness’. The correctness of these two heuristics is debatable. It is reasonable to think that many people would agree with the statements and that in several situations these heuristics might produce good results. Nonetheless, with more careful consideration, driven by System 2, people may support the implementation of automatic nudges. For instance, support for the use of automatic nudges may increase if information is provided about the fact that choices are often driven by uncontrolled factors, or which alternatives and options are available to promote positive policy goals. Finally, the likelihood of activating these heuristics seems to be influenced by the information that subjects are given, and it is much more likely when issues of deliberation and manipulation are salient in the description of the nudge. Likewise, these heuristics might be more likely to play a role when the domain is closer to a sensitive issue or moral area which would explain the outraged reactions to automatic nudges in domains such as organ donation or voting behaviour.

If preferences for automatic nudges are not the result of rigid moral considerations, but instead are influenced by the way in which nudges are presented, highlighting the mechanisms by which nudges work should decrease the acceptability of automatic nudges, whereas highlighting their effectiveness should improve it. To explore this issue, I investigate the following hypothesis:

*Hypothesis 1. Highlighting the mechanism by which nudges work reduces the acceptability of automatic nudges.*

*Hypothesis 2. Highlighting the effectiveness of nudges improves the acceptability of automatic nudges.*

If the reasoning on the acceptability of automatic nudges is mostly driven by System 1, the effect of this information should be more substantial when the judgment is made under time pressure.

*Hypothesis 3. The effect of information on the acceptability of automatic nudges increases when subjects evaluate this information under time pressure.*

The fact that the way in which nudges are explained and communicated might impact their acceptability is critical to understanding why they are approved or rejected and how to communicate their use to ensure acceptability.

- *Research question 2. How does the context of choice affect the acceptability of automatic nudges?*

The way in which automatic nudges are communicated is essential, but it is certainly not the only issue that affects their acceptability. Whether it is acceptable to affect people's choices triggering non-deliberative decision-making processes also depends on the aim of the nudge and the context in which it is applied. As argued in Chapter 2, the aim of a nudge is one of the factors that most affects people's opinions on their implementation: when people agree with the aim, they support nudges, when they oppose the aim, they oppose them. However, in the case of automatic nudges, the effect of the aim is not always conclusive. While supporting the aim is relevant, in some cases, it is not enough to support the implementation of automatic nudges.

Previous studies indicate that attitudes towards automatic nudges are also affected by the context in which they are applied. The effect of the context is relevant in two ways. Firstly, when judging the acceptability of automatic nudges, people take the domain in which nudges are applied into account. Studies indicate that disapproval of automatic nudges is higher in contexts of high stakes, such as in decisions that involve economic losses or particularly sensitive topics. In specific domains, even when people agree with the objective of the nudge, they reject the use of automatic nudges because they attach value to making conscious decisions<sup>19</sup>. The second crucial factor is the nature of the decision that the nudge is targeting. The acceptability of automatic nudges seems to be affected by the way in which particular choices are made in different contexts. In areas in which people perceived that their choices are influenced by the environment and outside their control, people appear to be more accepting of automatic nudges (Felsen et al., 2013).

The inconsistency on the acceptability of automatic nudges across domains and decisions indicates that people not only take the nudge into consideration when valuing nudges, they also consider the context in which the nudge is applied. Previous studies suggest that all of these factors are relevant. However, there are no studies that investigate how they relate to one another and which of these factors, if any, outweighs the others. From previous studies, it appears that people are more likely to accept (reject) the use of automatic nudges when they agree (disagree) with the aim of the nudge. People are more likely to accept (reject) the use of automatic nudges when the issue at stake is low stake (high stake). People

---

<sup>19</sup>Studies find general support for interventions to promote organ donation yet overall disapproval when it comes to using automatic nudges to promote this goal. Likewise, people tend to reject automatic nudges that involve economic losses or that target decisions related to political and religious values

are more likely to accept (reject) the use of automatic nudges when the nature of decisions is unconscious (conscious). However, it is unclear how these factors balance each other out. As a result, in order to investigate public support for nudges, it is essential to understand how people react to automatic nudges by taking the interaction between these factors into account. It is crucial to study them together, so as to isolate how a change in one of these factors affects the acceptability of nudges when the others remain fixed.

Study B explores the impact on the acceptability of nudges resulting from highlighting that nudges are used to change deliberate or non-deliberate behaviour and whether this effect varies depending on the issue at stake. Following findings in previous studies, I expect to see more support for automatic nudges when they change non-deliberative behaviour and when they do so in a low stake context. To explore these issues, I investigate the following hypothesis:

*Hypothesis 4. Automatic nudges are more accepted (rejected) when used to change non-deliberative (conscious/deliberate) choices.*

*Hypothesis 5. Automatic nudges are more accepted (rejected) in a low stake (high stake) context.*

*Hypothesis 6. The acceptability of automatic nudges increases (decreases) when they are used in low stake (high stake) contexts where decisions are unconscious (conscious).*

### **3. Study A. Are preferences for automatic nudges affected by how nudges are described?**

Study A analysed whether preferences for automatic nudges are affected by the available information and the time available to subjects evaluate the nudges. It used a survey experiment ( $n=314$ ) with a between-subject design that manipulates two factors: (i) the type of information about the interventions and (ii) the time available to subjects to review them and express how they feel about their implementations.

#### **3.1. Materials and Methods**

- *Sampling and recruitment*

The recruitment, sampling and survey were performed by Netquest, an international market research company with an IOS certified panel, with whom I worked to monitor each step of the planning, sampling and implementation of the survey. Respondents conducted the survey online using the *Nicequest* app, which works on different devices, including tablets, computers and smartphones. The survey was conducted in May 2019. Following Netquest's regular compensation practices, participants were rewarded with Netquest points (*Korus*) that they could later exchange within the same app for gifts or to make donations to a charity of their choice. Participants who took part in the national survey on nudges reported in Chapter 6 could not participate.

- *Study design and procedure*

The core of the study was a questionnaire about eight different measures to reduce the consumption of sugar-sweetened beverages (SSBs). Participants in all conditions were first presented with an introductory text explaining the increasing burden of obesity around the world due to increasing consumption of SSBs. They were then presented with a brief description of eight different measures to reduce the consumption of SSBs.

I chose to explore attitudes towards nudges in the healthy eating domain for several reasons. Firstly, it is a pro-self-domain but with pro-social consequences, because it contributes to reducing negative externalities. This implies that it is possible to assess how people feel about measures that can promote their wellbeing and the wellbeing of others. Secondly, both automatic and reflective nudges have been applied to this domain, making it possible to explore attitudes towards different measures. Finally, studies examining the acceptability of nudges in the healthy eating domain are frequent, which allows comparison.

Adapting material from Petrescu et al. (2016), participants read about three automatic nudges (size, location and default), three reflective nudges (labels, campaign and app) and two no-nudge interventions (tax and subliminal messages). As detailed in Table 10, the descriptions were short and straightforward and did not use the term ‘nudge’. The *administration* was cited as the nudger, avoiding possible political connotations associated with parties or regional and national governments. The different interventions were evaluated separately and appear in random order (See the full survey in Table 17, Appendix B).

**Table 10. Interventions included in the survey of Study A**

<b>Size</b>	Reducing the sugar-sweetened beverages containers to smaller sizes (price adjustment to new size).
<b>Location</b>	Controlling the positioning of sugary drinks in supermarkets and restaurants so that they are in places that are less visible and difficult to access.
<b>Default</b>	Establishing water as the drink included on restaurant menus by default.
<b>Labels</b>	Introducing traffic light labels on the packaging of different products and highlighting high sugar beverages in red to warn that they are unhealthy.
<b>Campaign</b>	Campaign to raise awareness about the problems associated with the consumption of sugary drinks and introduce messages and posters in shopping areas that provide information about the risks associated with the consumption of these products.
<b>App</b>	App that allows users to control their consumption of sugary drinks, notify them in case of excessive consumption and recommend alternative products.
<b>Tax</b>	Introducing a tax on sugary drinks to increase their purchase price.
<b>Subliminal messages</b>	Playing subliminal messages in movie and television commercials to induce the consumption of sugar-free and healthier drinks.

The study controlled (i) the information about the interventions (treatment 1) and (ii) the time participants had to evaluate them (treatment 2). *Treatment 1. Information* was designed to test whether information about the effectiveness of the measures and the mechanism by which the measures work affected people’s opinions of the different interventions. The descriptions under the *effectiveness conditions* highlighted that the interventions were effective in reducing the consumptions of sugar-sweetened beverages. The descriptions of the measures under the *mechanism conditions* highlighted whether interventions changed behaviour through conscious or non-conscious mechanisms. In the case of automatic nudges and *subliminal messages*, participants were informed that the interventions affect behaviour in an “automatic and subconscious way”, whereas in the case of reflective nudges and *tax*, participants were told that the interventions changed behaviour in “a conscious way”. The control condition included the simple definitions detailed in Table 10. *Treatment 2. Time pressure* was designed to inflict cognitive load to the evaluation task. Under the *time-pressure conditions*, participants had forty seconds to read the descriptions and eight seconds to answer the main question. They were informed about these limits at the beginning of the study and had a visible timer on-screen. In the *non-time pressure conditions*, participants did not have any time restrictions. The experiment tested six conditions with a combination of treatments. Table 11 details them and indicates the sample for each condition.

**Table 11. Summary of treatments and experimental conditions of Study A**

		Information about the measures		
		No information	Effectiveness	Mechanism
Time pressure	Time pressure	C1 (n=56)	C2 (n=46)	C4 (n=47)
	No-time pressure	C0 (n=54)	C3 (n=58)	C5 (n=53)

- *Response variables*

Following the description of each measure, participants were asked: ‘Do you agree with the implementation of this measure?’ Each response was rated on a 5-point-Likert scale ranging from 1 ('Not at all') to 5 ('Yes, totally'). Likewise, the study included an open question that asked participants what they considered most relevant when assessing the acceptability of different measures. The question was included to capture what people had in mind when evaluating the nudges, as well as to measure the effectiveness of the treatments under conditions C2, C3, C4 and C5.

The study also asked participants about their perceived effectiveness and their perceived manipulation for each measure. Under the no-information conditions (C0 and C1), subjects were asked about both; firstly, whether they thought that the interventions were going to be effective, and then, whether they

perceived that the measure manipulated behaviour. Under the effectiveness conditions (C2, C3), they were only asked about perceived manipulation and, under the mechanism conditions (C4, C5), they were only asked about perceived effectiveness. The responses were assessed after the general survey and ranked on a 5-point-Likert scale ranging from 1 ('No at all') to 5 ('Yes, totally').

- *Co-variables questionnaire*

In addition to the main survey, the study included an extensive questionnaire on complementary predictors of acceptability. Participants answered the questionnaire once they had completed the survey. The questionnaire collected sociodemographic variables, political values and information on individual behaviour and individual dispositions.

Firstly, to consider the effect of sociodemographic characteristics, the study collected gender, age, activity, education and nationality.

Secondly, the questionnaire asked respondents about political orientation, trust in institutions and attitudes towards government interfering for individual good. Using Spanish Centre for Sociological Research (CIS) methodology, ideology was measured by asking respondents to self-identify on a scale from 0 to 10, where 0 corresponded to the extreme left and 10 to the extreme right. Respondents were also asked if they trust public institutions, whether they favoured government interference to promote healthy eating, and whether the government should limit individual decisions to improve individual wellbeing or should not interfere in people's decisions even if their choices have negative consequences for themselves. Each of these responses was rated on a 5-point-Likert scale ranging from 1 ('Not at all') to 5 ('Yes, totally').

Thirdly, four responses were used to assess diet-related health consciousness: self-perception of having a healthy diet ('Do you maintain a healthy diet?'), importance of maintaining a healthy diet ('Is it important for you to maintain a healthy diet?'), beliefs about the impact of consumption of SSBs on health ('Do you think your health is affected by the consumption of SSBs?'), and degree of attention to sugar in one's diet ('Do you usually look at the sugar content of the products you consume'). Each of these responses was rated on a 5-point-Likert scale ranging from 1 ('Not at all') to 5 ('Yes, totally').

The questionnaire also included three questions to assess dietary status. Participants were asked how often they consumed sugary drinks (answered on a response scale with six categories: more than twice a day, once a day, 2 to 4 times a week, once a week, occasionally - less than once a week, rarely or never); and whether they had been on a diet during the past 12 months (answered on a binary response scale: yes or no). The questionnaire also asked participants for their height and weight in order to calculate their BMI.

Finally, the questionnaire also included five questions to collect individual beliefs on the factors that drive behaviour and obesity. Two questions assessed the perceived need for help in making healthier

choices for oneself and others ('Do you consider that in day-to-day life you/others would benefit from getting some kind of help to have a more balanced and healthy diet?'). Responses were rated on a 5-point-Likert scale ranging from 1 ('No, not at all') to 5 ('Yes, totally'). Likewise, two questions assessed participants' perceptions of situationist and dispositionist tendencies ('Do you consider that your/others' behaviour is mainly determined by choices and decisions or the context and circumstances that surround me/them?'). Responses were rated on a 5-point-Likert scale ranging from 1 ('Because of my/their choices and decisions') to 5 ('Because of the context and circumstances that surround me/others'). A question was included to collect beliefs regarding the causes of obesity ('Why do you think people are overweight and obese?'). Participants answered on a 5-point-Likert scale ranging from 1 ('Because there are too many opportunities to access unhealthy products and difficulties to lead a healthy life') to 5 ('Because people have self-control problems and do not correct their behaviour).

All of the variables included in the post-experiment questionnaire are incorporated in the analysis as potential predictors of acceptability (See the questionnaire in Table 17, Appendix B).

- *Statistical analysis*

324 participants conducted the survey; four participants were excluded from the final sample due to inattentive responses, and six were excluded due to missing values on the response variables. The final sample consisted of 314 participants. Their average age was 48 (S.D. = 16), 50% were female, 47% were occupied and working full time and 47% had high educational attainment (See characteristics of the participants in the study in Table 18, Appendix B).

The analysis uses the acceptability of each nudge as the primary outcome measure. Likert-type responses were analysed as categorical variables. The relationship between variables was first explored with chi-square test and then with logistic regression. Given the distribution of the data, responses were recoded (See information on data preparation in Appendix B. Section 1.3). To perform the logistic regression, the response variables were dichotomised. Since the primary objective was to explore whether treatments affect acceptability, the answers were recoded into two categories: a category denoting acceptability (when 'strongly agree' and 'agree' responses are combined) and a category denoting no-acceptability (including 'neutral', 'disagree' and 'strongly disagree' responses). Multicollinearity was controlled using the variance inflation factor. Coding and analysis were performed using R.

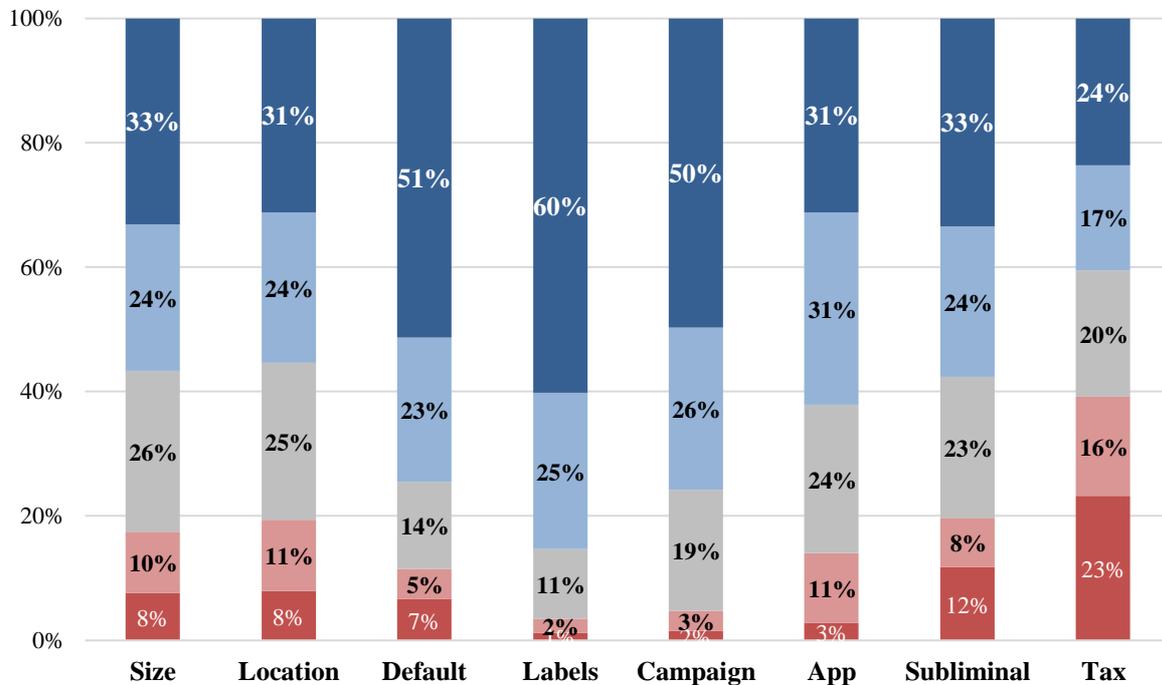
## **3.2. Results**

### *3.2.1. Acceptability of the interventions*

Figure 4 shows the percentage of responses for each measure. Participants approved of all the interventions with high and moderate levels of support, except for tax. The use of traffic-light labels on the packaging of SSBs was the most popular measure, with 60% of respondents strongly agreeing with

their implementation and less than 3% of respondents disagreeing with their use. By contrast, the implementation of tax on SSBs was the most controversial measure, where 41% supported it, but 39% opposed it.

**Figure 4. Acceptability of the eight interventions by response.**



Percentage of responses by category for each intervention. n= 314. Categories: ‘Strongly agree’ (5) in dark blue, ‘Agree’ (4) in light blue, ‘Neutral’ (3) in grey, ‘Disagree’ (2) in pink and ‘Strongly disagree’ (1) in red. Automatic nudges: size, location and default. Reflective nudges: labels, campaign and app. No-nudge measures: subliminal messages and tax.

Both automatic and reflective nudges were supported. As expected, reflective nudges were more accepted than automatic nudges. The acceptability of reflective nudges and automatic nudges produced strikingly similar patterns of results, but with two exceptions. Two reflective nudges (labels and campaign) were highly supported, and less than 5% of respondents disapproved of their use. These results are consistent with previous studies that have found high approval for these interventions, particularly when used to promote healthy choices (Evers et al., 2018; Hagmann et al., 2018; Sunstein, 2017b). Two automatic nudges (size and location) also have almost identical patterns of responses. The majority of respondents supported the use of smaller SSBs containers and the positioning of these products in difficult to access areas (57% and 55% respectively, when ‘strongly agree’ and ‘agree’ responses are combined). However, both interventions had a lower acceptability percentage and more detractors than the two reflective nudges (18% and 19% respectively, when ‘strongly disagree’ and ‘disagree’ responses are combined). Again, the results are consistent with previous studies that found moderate acceptability for these type of interventions in the healthy eating domain (Evers et al., 2018; Hagmann et al., 2018; Petrescu et al., 2016)

The default nudge (automatic) and the app nudge (reflective) presented a distribution of responses that resembles the structure of the contrary type of nudge. The measure to establish water as the default

drink included on restaurant menus was well supported with 51% of respondents strongly agreeing with its implementation. While the implementation of the default had more detractors than the two reflective nudges (12% when 'strongly disagree' and 'disagree' responses are combined), the percentage of 'strongly agree' responses matches the rates for the labels and the campaign. By contrast, the use of an app to self-regulate the consumption of SSBs had moderate levels of support and more detractors, matching the pattern of responses for automatic nudges. Some factors could explain the lower acceptability for the app. It could be that respondents dislike self-monitoring their behaviour, have concerns about privacy or distrust their own determination to use it. The high acceptability of the default is more understandable. As found in Chapter 6 and other studies, people seem to recognise the power of defaults. People favour their implementation when they are used to promote positive outcomes, and in a domain in which they agree with their use. Respondents might also perceive this as an easy to opt-out default, which will increase its acceptability. The results suggest that negative attitudes towards defaults are not only motivated by how they affect behaviour but also by what their aim is.

The three most approved of interventions, the labels, the campaign and the default, were also the three interventions perceived as being more effective and less manipulative. The two remaining automatic nudges (size and location) and the remaining reflective nudge (app) were not perceived as being effective but nor were they seen as being manipulative, with a majority of respondents in the neutral or disagree categories (See data on perceived effectiveness and perceived manipulation in Figures 9 and 10, Appendix B).

The results on the acceptability of the six nudges reveal that, in general, people preferred reflective nudges to automatic nudges. These results were expected and are consistent with previous research that suggests that people have concerns about transparency and prefer interventions that engage their reflective capacities. However, the responses for the default and the app broke the pattern, indicating that opposition to automatic nudges is not categorical and that active engagement or consciousness is not the only important criteria when evaluating the acceptability of nudge-related measures.

Support for subliminal messages produced surprising results. It had a distribution of respondents similar to the responses for size and location; 57% of respondents approved of the measure (when 'strongly agree' and 'agree' responses are combined) and 20% of respondents disapproved of the intervention (when 'strongly disagree' and 'disagree' responses are combined). It was surprising to find moderate levels of support rather than strong opposition for subliminal messages. However, previous studies on nudge acceptability have also found moderate support for subliminal messages used against overeating (Sunstein, 2017b) with there being even high support in some countries (Sunstein et al., 2018). Different factors could explain the support for the measure. Firstly, it is possible that people strongly agree with the policy goal and accept the intrusiveness of the measure to favour the outcome. Secondly, it is possible that if they are included between diverse interventions and presented as an option that the administration is thinking about implementing, subliminal messages appear to be a valid tool for

influencing people's choices. People did not perceive this intervention as being particularly effective (44% thought it would be effective, but 24% thought it would not). Likewise, 52% thought that use of subliminal messages was a manipulative intervention (when 'strongly agree' and 'agree' responses are combined), but 19% thought it was not (when 'strongly disagree' and 'disagree' responses are combined). These results suggested that people could be underestimating the effect that subliminal messages might have on their own behaviour and the behaviour of others, which could also explain why opposition was not higher.

The implementation of a tax on SSBs was the most disapproved of measure. The result is consistent with previous studies which have found low acceptability for taxes in the healthy eating domain (Evers et al., 2018; Petrescu et al., 2016). The tax was also the measure perceived as being the most ineffective, and the second most manipulative.

Looking at the distribution of responses in the perceived effectiveness and perceived manipulation questions, it appears that people were unsure what to answer. These questions had a higher percentage of neutral responses (between 23% and 31% in the effectiveness questions, and between 28% and 38% in the manipulation questions). Accordingly, it appears that people had unclear and, in some cases, inaccurate beliefs about both issues. It looks as though people's responses about effectiveness and manipulation reflected their preferences about the measures, denoting a strong association between liking an intervention and seeing it as effective and non-manipulative.

To sum up, results were very consistent with previous studies. People accepted the use of nudges to reduce the consumption of SSBs, indicating that people's attitudes towards nudges are positive in domains in which they agree with their aims. Informative and educative nudges (labels and campaigns) were preferred, and taxation was opposed. Although automatic nudges (size and location) were approved of by the majority, there was less acceptance of them than other measures and greater opposition to them. Interestingly, the implementation of a default healthy drink was highly supported, whereas the use of a self-monitoring app gathered only moderate support. The results showed that, in general, people preferred informative and educative interventions, although they also approved of the use of automatic nudges.

### *3.2.2. Effect of the experimental treatments*

Logistic regression analyses were conducted to explore the influence of the experimental conditions on the acceptability of the eight measures. As can be seen in Table 12, the effect of the experimental condition was not significant for any nudge or the subliminal messages. However, it did affect the acceptability of the tax.

*Treatment 1. Information* was ineffective when it came to changing acceptability levels for the six nudges and for the measure that included subliminal messages. Highlighting that the measures were effective did not increase their acceptability. Likewise, highlighting the mechanism through which the

interventions work (either as a conscious or non-conscious mechanism), did not have any impact on the acceptability levels either. As such, I found no evidence to support hypothesis 1 and 2.

**Table 12. Logistic regression models to assess the impact of conditions on acceptability.**

	Size	Location	Default	Labels	Campaign	App	Subliminal	Tax
<b>Intercept</b>	-0.53(0.54)	-0.09(0.53)	1.49(0.61)	1.68(0.76)	0.82(0.62)	0.75(0.55)	-0.14(0.54)	-1.69(0.58)**
<b>Time pressure</b>	0.12(0.39)	-0.55(0.39)	-0.23(0.43)	-0.66(0.57)	-0.46(0.46)	-0.36(0.40)	-0.58(0.40)	0.92(0.43)*
<b>Effectiveness + tp</b>	0.44(0.41)	0.20(0.41)	0.32(0.49)	0.63(0.74)	-0.36(0.48)	-0.17(0.43)	-0.22(0.42)	1.32(0.45)**
<b>Effectiveness</b>	0.19(0.38)	-0.14(0.38)	0.30(0.45)	-0.53(0.56)	-0.21(0.46)	-0.28(0.40)	-0.36(0.40)	0.78(0.43) .
<b>Mechanism + tp</b>	0.67(0.41)	0.21(0.41)	0.06(0.46)	-0.70(0.58)	-0.05(0.50)	-0.41(0.42)	-0.64(0.42)	1.07(0.44)*
<b>Mechanism</b>	0.74(0.40) .	0.06(0.40)	-0.08(0.44)	-0.43(0.58)	-0.15(0.48)	-0.04(0.41)	-0.04(0.41)	1.23(0.43)**
<b>Gender (F)</b>	0.41(0.25) .	0.17(0.25)	0.10(0.29)	-0.18(0.35)	0.08(0.29)	-0.07(0.26)	0.50(0.26)*	0.16(0.26)
<b>Age</b>	0.01(0.01)	0.01(0.01)	-0.01(0.01)	0.01(0.01)	0.01(0.01)	0.00(0.01)	0.01(0.01)	0.01(0.01)

The control condition taken as reference. Probability of 1 (agreeing) over 0 (neutral or disagreeing). Gender and Age taken as control variables. Standard errors in parentheses. Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

*Treatment 2. Time pressure* did not perform as intended. On average, people finished the survey in 10.39 minutes (SD=6.11); however, the time taken was higher under the time pressure conditions. The results from a one-way analysis of variance (ANOVA) revealed that the difference was significant for the pairs: C4 - C0 (diff: 3.50\*) C5 – C1 (diff: -3.32\*), C4-C3 (diff: 3.47\*) and C5-C4 (diff: -4.07\*\*)20. Altering the time that subjects had to read and answer the questions was ineffective when it came to inducing a sensation of time pressure. Respondents, instead, took more time to do the survey, particularly in condition 4 (mechanism + time pressure). I did not anticipate this problem because the time pressure treatment was tested in a pre-test and worked as intended. It is difficult to assess what caused the problem. The one-way ANOVA analysis revealed that duration time was unrelated to age and educational attainment. Gender was significant, with women taking more time than men (diff: -1.47 \*), yet the effect of gender did not override the effect of the treatment. Perhaps, the on-screen timer acted as a signal of the time participants *should* use to read the questions21. Regardless, the problems with the time pressure treatment make it difficult to interpret its effects and test hypothesis 3.

Treatments did affect the acceptability of taxation. Information had a positive effect. Highlighting that the measure was effective in reducing the consumption of SSBs increased the acceptability of the tax. The effect of the information on effectiveness was stronger in the time pressure condition. Likewise, highlighting that the measure influenced behaviour consciously also had a positive impact on its

20Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

21The time pressure conditions might have produced more attentive responses due to a possible signalling effect of the timer and because subjects took longer to answer. If this is the case, these treatments might have produced more utilitarian judgments and, therefore, increased the acceptability of the interventions.

acceptability; the effect was similar in the time pressure and no time pressure conditions. By itself, time pressure also had a positive effect on the acceptability of taxes. This could be because respondents used more time to read and deliberate about the interventions, thereby favouring a utilitarian judgment.

To further explore the effect of information, I drop the time pressure treatment and combined experimental groups according to the type of information they provided. C0 and C1 became *the control condition* (no information,  $n=110$ ), C2 and C3 the *effectiveness condition* ( $n=104$ ), and T4 and T5 the *mechanism condition* ( $n=100$ ). Logistic regression was performed. The analysis yielded very similar results. Information only influenced the acceptability of the tax. Both, information about effectiveness and about the mechanism increasing the acceptability of taxation.

### 3.2.3. Predictors of acceptability

Logistic regression analyses were conducted to explore the influence of socioeconomic and individual differences on attitudes towards the eight measures. Table 13 shows the results. Treatment 2 (time pressure) was not included. The six initial conditions were combined into three, based on the type of information they provided. Some variables were excluded to avoid problems of multicollinearity (See information on data preparation in Section 1.3, Appendix B).

The logistic regression analyses revealed that different factors affect the acceptability of the policies; however, none of the predictors systematically affected attitudes across all the interventions. Sociodemographic variables had little association with acceptability. Gender affected the support for two automatic nudges (size and location) with women more likely to approve both interventions. The result is consistent with previous studies and with the results in Chapter 6. Women tend to be more in favour of interventions to promote healthy eating because they are more concerned with eating healthily and because they tend to be more empathetic and more concerned with others' wellbeing than men (Evers et al., 2018; Hagmann et al., 2018). Therefore, it makes sense that the support for these two automatic nudges is higher among women. Age affected the acceptability of the default measure, with older people less in favour of its implementation. The result ties up with previous studies that found older people are less in favour of intrusive interventions (Sunstein et al., 2018).

Political factors had modest effects and did not generally influence the acceptability of the eight policies. As expected, ideology did not have a systematic effect on the acceptability of the measures. Those that self-identified as being on the right were less likely to accept the implementation of the app to self-control the consumption of SSBs, possibly because of worries associated with privacy or data protection. The minimal effect of ideology aligns with previous research and with the findings reported in Chapter 6. Trust in institutions also had a very minimal effect and only influenced the acceptability of the size measure, with those that trust institutions being more in favour of its implementation. I expected trust in institutions to have a higher correlation to acceptability, in particular to the acceptability of automatic nudges. The results contradict these expectations, and I found no evidence

to support them. It is, however, worth noticing that while a large majority of respondents declared that they do not trust public institutions (46% when ‘strongly disagree’ and ‘disagree’ responses are combined), this factor did not undermine the acceptability of the eight policies.

Attitudes towards government interference were also generally unrelated to acceptability. Support for government interference to promote healthy eating was generally high but only significantly related to the implementation of the campaign. Attitudes towards the role of public administrations in individual choices were significantly related to the support of the tax. Those who consider that the public administration should not interfere in individual choices, even when choices have negative consequences for individual wellbeing, were less likely to approve of the use of a tax. I found these results interesting, because it appears that the respondents perceived the tax to be the most intrusive intervention, and not automatic nudges. Overall, I found no evidence to suggest that trust in institutions and attitudes towards government interference affect the acceptability of the measures. The possibility remains that these factors are more relevant when the aim of the nudges is more controversial or more politically contested.

**Table 13. Logistic regression models predicting the acceptability for the eight interventions.**

	Size	Location	Default	Labels	Campaign	App	Subliminal	Tax
Intercept	-2.35(1.78)	-1.54(1.79)	-0.48(1.92)	-2.62(2.38)	-2.74(2.09)	-0.98(1.66)	1.9(1.64)	-1.61(1.8)
Effectiveness + TP	0.27(0.41)	0.36(0.4)	-0.27(0.48)	0.42(0.6)	-0.24(0.53)	-0.46(0.41)	-0.09(0.39)	0.17(0.41)
Mechanism + TP	0.81(0.43).	0.64(0.41)	-0.66(0.47)	-0.39(0.56)	0.3(0.53)	-0.31(0.41)	-0.06(0.4)	0.3(0.42)
Gender (F)	1.03(0.41)*	0.77(0.39)*	0.09(0.46)	0.06(0.58)	0.53(0.49)	0.07(0.39)	0.48(0.37)	-0.07(0.39)
Age	0(0.01)	-0.01(0.01)	-0.03(0.01)*	0(0.02)	0.02(0.02)	0.01(0.01)	0.01(0.01)	-0.01(0.01)
Secondary education	0.35(0.59)	-0.38(0.6)	1.06(0.64) .	0.09(0.79)	-0.48(0.71)	-0.06(0.59)	-0.24(0.6)	-0.37(0.61)
University education	0.24(0.58)	-0.39(0.58)	0.57(0.62)	0.5(0.81)	-0.04(0.71)	0.39(0.57)	-0.62(0.58)	0.11(0.58)
Ideology (Centre)	0.74(0.39).	0.07(0.38)	-0.44(0.45)	0.45(0.5)	-0.51(0.47)	-0.27(0.38)	-0.42(0.36)	-0.3(0.38)
Ideology (Right)	0.52(0.51)	0.05(0.48)	0.02(0.56)	1.38(0.72) .	-0.61(0.57)	-1.17(0.48)*	-0.54(0.46)	-0.06(0.5)
Trust in institutions (n)	0.09(0.38)	-0.05(0.37)	-0.46(0.44)	-0.2(0.51)	0(0.45)	0.08(0.38)	0.46(0.37)	0.68(0.39) .
Trust in institutions	1.19(0.51)*	0.57(0.48)	0.1(0.56)	0.72(0.72)	1.29(0.71) .	0.46(0.49)	-0.13(0.44)	0.58(0.48)
Interference to promote healthy eating (n)	-0.73(0.91)	0.35(0.9)	-0.43(0.86)	0.84(1.01)	1.13(0.94)	0.07(0.86)	-0.22(0.83)	-0.3(0.97)
Interference to promote healthy eating (s)	0.52(0.89)	0.33(0.89)	0.91(0.89)	1.58(1.04)	2.92(1.01)**	1.23(0.87)	0.33(0.82)	-0.51(0.96)
Interference to promote healthy eating (ss)	1.07(0.92)	0.11(0.91)	0.15(0.94)	1.79(1.09)	2.89(1.02)**	1.13(0.9)	0.14(0.86)	-0.72(0.97)
Interfer. to promote pro-self-outcomes (n)	-0.32(0.42)	-0.67(0.41)	-0.23(0.5)	-0.73(0.57)	0.18(0.51)	-0.06(0.41)	0.07(0.39)	-1.08(0.41)**
Interfer. to promote pro-self-outcomes (h)	-0.38(0.47)	-0.89(0.46).	-0.52(0.54)	0.04(0.69)	1.14(0.62) .	0.72(0.47)	-0.02(0.43)	-1.55(0.47)**
Importance of healthy diet (n)	0.68(0.54)	0.35(0.52)	0.43(0.58)	0.83(0.62)	1.39(0.59)*	0.52(0.5)	-0.22(0.52)	-0.49(0.56)
Importance of healthy diet (s)	0.22(0.57)	0.89(0.57)	0.92(0.64)	1.31(0.7) .	0.77(0.63)	0.87(0.54)	-0.27(0.56)	-0.02(0.61)
Beliefs on SSBs impact (n)	0.56(0.5)	-0.16(0.49)	-0.28(0.53)	-0.62(0.62)	-0.54(0.59)	0.16(0.5)	-0.14(0.47)	-0.14(0.51)
Beliefs on SSBs impact (h)	1.57(0.41)***	0.35(0.39)	0.23(0.47)	0.8(0.55)	0.76(0.49)	-0.01(0.38)	-0.02(0.37)	0.05(0.39)
Attention to sugar (n)	1.3(0.6)*	0.05(0.54)	-0.18(0.61)	0.83(0.77)	-0.14(0.63)	-0.26(0.51)	-0.22(0.5)	-0.42(0.58)
Attention to sugar (h)	-0.28(0.44)	0.28(0.43)	0.26(0.5)	-0.07(0.58)	-0.12(0.52)	0.66(0.43)	0.1(0.41)	0.81(0.44) .

Probability of 1 (agreeing) over 0 (neutral or disagreeing). Standard errors in parentheses. All categorical predictors with the exception of Age and BMI included as continuous variables.

Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' . (n) = neutral. (s) = support. (ss) = strong support. (h)=high.

	<b>Size</b>	<b>Location</b>	<b>Default</b>	<b>Labels</b>	<b>Campaign</b>	<b>App</b>	<b>Subliminal</b>	<b>Tax</b>
Sugar consumption (regular)	-1.66(0.72)*	-0.24(0.67)	-0.2(0.7)	-0.86(0.92)	-0.74(0.83)	0.31(0.66)	-1.44(0.68)*	-0.52(0.68)
Sugar consumption (once a week)	-1.22(0.74)	-0.83(0.7)	0.29(0.76)	-0.21(0.98)	-0.91(0.87)	0.04(0.69)	-0.96(0.71)	-0.07(0.71)
Sugar consumption (occasional)	-0.68(0.66)	-0.63(0.62)	0.09(0.67)	-0.66(0.91)	-0.31(0.82)	-0.23(0.61)	-1.6(0.65)*	0.73(0.61)
Sugar consumption (never)	-0.26(0.69)	-0.36(0.66)	0.53(0.72)	-0.52(0.94)	-1.17(0.83)	-1.09(0.65)	-1.35(0.67)*	1.25(0.65)
Dieting	0.35(0.42)	0.68(0.42)	0.63(0.48)	1.15(0.58)*	0.82(0.5)	0.46(0.4)	1.01(0.39)**	0.41(0.41)
Self-perception of need for external help to maintain a healthy diet (n)	-0.12(0.47)	-0.20(0.46)	0.2(0.52)	-0.32(0.6)	-0.69(0.6)	0.04(0.46)	0.13(0.44)	-0.4(0.48)
Self-perception of the need for external help to maintain a healthy diet (n)	-0.24(0.48)	-0.35(0.48)	0.42(0.54)	0.18(0.64)	-0.81(0.61)	-0.06(0.46)	0.48(0.43)	-0.14(0.48)
Self-perception as dispositionists	0.36(0.5)	0.36(0.49)	-0.32(0.58)	-1.13(0.72)	0.26(0.63)	-0.41(0.48)	-0.43(0.46)	0.15(0.51)
Self-perception as situationist (n)	0.05(0.47)	0.86(0.48)	-0.24(0.53)	-1(0.69)	-0.33(0.57)	-0.17(0.45)	0.18(0.43)	0.9(0.48)
Self-perception as situationist	0.37(0.5)	-0.09(0.49)	-0.79(0.55)	-1.23(0.7)	-0.69(0.6)	-0.39(0.48)	0.16(0.47)	0.81(0.5)
Perception of others as dispositionists	-1.03(0.64)	-1.3(0.59)*	-0.04(0.64)	0.06(0.88)	-0.46(0.74)	-0.17(0.58)	-0.52(0.56)	-0.67(0.62)
Perception of others as situationist(n)	-0.45(0.61)	0.19(0.56)	1.12(0.6)	0.13(0.82)	-0.12(0.71)	0.36(0.54)	-0.45(0.52)	0.44(0.57)
Perception of other as situationist	-1.28(0.66)	-0.87(0.61)	0.05(0.63)	-0.81(0.85)	-0.15(0.77)	0.23(0.61)	-0.44(0.58)	-0.26(0.62)
Obesity caused by lack of willpower (n)	0.33(0.49)	0.06(0.49)	0.46(0.59)	-0.12(0.64)	0.62(0.58)	1.42(0.5)**	-0.37(0.46)	0.14(0.51)
Obesity caused by lack of willpower (h)	0.57(0.38)	-1.04(0.38)**	-0.56(0.44)	-0.29(0.53)	0.31(0.48)	-0.04(0.36)	-0.69(0.35)*	-0.34(0.37)
Perception of need for other external help to maintain a healthy diet (n)	-0.5(0.48)	0.05(0.46)	-0.11(0.55)	-0.17(0.62)	-0.42(0.58)	0.43(0.48)	-0.22(0.46)	0.61(0.5)
Perception of others need for external help to maintain a healthy diet (h)	-0.03(0.43)	1.27(0.45)**	0.04(0.51)	0.43(0.57)	0.1(0.54)	0.32(0.43)	0.16(0.42)	0.38(0.45)
BMI	0.03(0.04)	0.06(0.04)	0.07(0.05)	0.09(0.06)	0.01(0.05)	-0.04(0.04)	-0.02(0.04)	0.05(0.04)

Table 13 continuing. Significance codes: 0 '\*\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1.

Diet-related health consciousness variables were also generally unrelated to the acceptability of the eight interventions. The importance to maintain a healthy diet was related to support for the campaign measure. The belief that SSBs are harmful to health increased the support for the size measure. Likewise, the degree of attention to sugar in one's diet also influenced the acceptability levels of size, with those paying medium attention being more in favour of implementing the measure. It may be the case that people who do not always pay attention to the sugar content but do care about reducing its consumption support the size measures because it does not require them to be vigilant. Overall, diet-related health consciousness was weakly related to the acceptability of the eight measures.

Dietary status had a modest effect. Self-reported consumption of SSBs influenced the acceptability of size and subliminal messages. Those that self-reported regular consumption of SSBs were less likely to support the reduction of the container sizes of SSBs and the use of subliminal messages. These results tied up with previous findings and with the results reported in Chapter 6, which document lower support among people that engage in the behaviour targeted by the nudges. At the same time, those that self-reported minimal or no consumption of SSBs opposed the use of subliminal messages to induce the consumption of sugar-free and healthier drinks. The adverse reaction makes sense since the subliminal messages policy is the only one that can affect their behaviour even when they do not need to change their consumption habits. Dieting was also related to the acceptability of the traffic light labels and the use of subliminal messages. Those that self-reported being on a diet or that they had been on a diet during the past 12 months were more likely to support both interventions. Again, the result is consistent with previous findings that document higher support for nudges among those that have personal preferences that are in line with the nudges targeted aim (Evers et al., 2018; Hagmann et al., 2018). BMI was unrelated to the acceptability of all the measures. Overall, the effect of diet-related health consciousness and dietary status was weaker than expected. However, the acceptability of some interventions is higher among those who have individual preferences that align with their objectives but is lower among those that engage with the behaviour that the interventions target.

Subjective perceptions of how the environment influences behaviour and health-related choices did not have a relevant impact on the acceptability of the interventions. Perception about the personal need for external help to maintain a healthy diet and self-perception of dispositionism or situationism were both unrelated to the acceptability of all the measures. By contrast, people's perceptions about how other people make choices did influence the acceptability of some of the measures. Those that believe that people's choices are mostly dispositional were less likely to approve of the positioning of SSBs in less accessible areas. Possibly the perception that people act according to their own dispositions made this intervention appear ineffective and, therefore, people disapproved of its use. People's beliefs about the causes of being overweight and obesity affected their acceptability for two of the nudges included in the survey. Those that believe that being overweight and obesity are a combination of the influence of external factors and a lack of willpower were more likely to favour the self-regulation app. It may be

the case that, when people recognise that lack of willpower as part of the problem, they favour interventions that increase accountability and may help others to self-regulate their consumption. By contrast, attributing being overweight or obesity to a lack of willpower reduced support for changes in positioning of SSBs and the use of subliminal messages. It is unclear why. It may be the case that people saw these interventions as being too intrusive on the choices of those who do have control, but still ineffective for those that lack willpower. Finally, those that perceived that others need external help to maintain a healthy diet were more likely to support the change of location measure. It makes sense that people that view the environment as the main problem show more support for an intervention that requires a direct modification of the context of choice. Overall, perceptions of the way in which their environment influences general behaviour and health-related choices were less related to the acceptability of the measures than expected.

The predictors considered in the study had little association with the acceptability of the measures. However, the analysis did show the influence of some relevant factors. The study has some limitations that need to be taken into consideration when drawing general conclusions from these results. Firstly, the study has a small sample size. Secondly, the study loses statistical power as a result of dichotomising the primary outcome variable and some of the predictors. The possibility remains that the effect of the factors considered may be more relevant with a larger sample.

#### **4. Study B. How does the context of choice affect the acceptability of automatic nudges?**

Study B investigated whether the acceptability of automatic and reflective nudges varies depending on (a) the nature of the decision that the nudges are targeting and (b) the domain in which the nudges are applied. The experiment was conducted online ( $n= 228$ ) and used contrastive vignettes to explore the ways in which these two factors affect the acceptability of automatic nudges.

##### **4.1. Materials and Methods**

- *Sampling and recruitment*

Netquest performed the recruitment, the sampling and the survey. Respondents conducted the survey online using the *Nicequest* app, during July of 2019. Participants were rewarded with Netquest points (*Korus*). Those who participated in the national survey about nudges and on study A were excluded from participation.

- *Study design and procedure*

Participants were first introduced to a text informing them that they would read about a hypothetical scenario and were then randomly assigned to read one (and only one) of eight contrastive vignettes. The vignettes presented a scenario in which the public administration had implemented a nudge to either reduce the consumption of SSBs (scenario A) or encourage organ donation (scenario B) and described

how the measure had affected the behaviour of a third person (Martin). Two factors varied across the vignettes: the description of the nudge and its type of influence (treatment 1) and the description of Martin's character (treatment 2). The other aspects of the vignettes had minimal variations.

*Treatment 1. Influence of the nudge* manipulated the type of influence of the nudges, described either as conscious or unconscious. In the *conscious conditions*, participants were informed that the nudge had affected Martin's decisions in "an unconscious way, beyond his perception and control" and emphasised that "Martin did not perceive that the measure influenced his behaviour". In the *unconscious conditions*, participants were informed that the nudge "made Martin reflect and affected his decisions in a totally conscious way" and emphasised that "Martin did perceive that the measure influenced his behaviour". *Treatment 2. Type of behaviour* manipulated the description of Martin's behaviour and described him as a dispositionist or a situationist. In the *dispositionist conditions*, Martin was described as "a person who chooses according to his preferences, who likes to be in control and make his own decisions". In the *situationist conditions*, Martin was described as "a person who does not always pay attention to his surroundings" (See the vignettes in Tables 19 and 20, Appendix B).

The vignettes are about two representative policy scenarios: organ donation and healthy eating. The selection of these two scenarios allowed me to explore whether the hypothesis would hold across two domains with different aims (pro-self and pro-social) and different stakes (high and low). I consider healthy eating to be a low-stakes domain. According to previous research on public attitudes towards nudges, automatic nudges tend to be generally accepted when they are used to promote healthy eating. Likewise, choosing what to eat is a choice that we make several times a day and that it is easy to modify. By contrast, I considered organ donation to be a high-stakes decision. Organ donation is a sensitive choice that involves religious and moral issues and that we probably only take once in a lifetime. Likewise, previous studies indicate that people disapprove of the use of automatic nudges in this domain, which indicates that people have worries about autonomy and would like to be able to make their own choices regarding organ donation. The experiment tested four conditions in each scenario using a combination of treatments. Table 14 details them and indicates the sample for each condition.

**Table 14. Summary of treatments and the eight experimental conditions of Study B.**

		Type of behaviour			
		Situationist		Dispositionist	
Type of nudge	Automatic	AS_A (n=46)	AS_B(n=45)	AD_A(n=21)	AD_B(n=29)
	Reflective	RS_A(n=19)	RS_B(n=14)	RD_A(n=25)	RD_B(n=29)

Abbreviations stand for: AS (Automatic + Situationist); RS (Reflective + Situationist), AD (Automatic + Dispositionist) and RD (Reflective + Dispositionist). \_A (healthy eating scenario), \_B (organ donation scenario).

- *Response variables*

After reading the vignette, the respondents answered a set of questions designed to capture their attitudes about the different scenarios.

Firstly, two questions were included to gather respondents' general impression of the way in which Martin's behaviour was influenced. After reading the vignette and on the same screen, respondents were asked a closed-ended question: 'Considering what you know about Martin, how do you feel about influencing his decisions in this way?' The question was answered on a continuous scale ranging from 0 to ('Very bad') to 100 ('Very good'). Then, they were also asked an open-ended question: 'Can you briefly indicate what motivated your answer in the previous question?' Subjects had to answer using a maximum of 300 characters.

Secondly, a set of five questions collected respondents' perceptions of the nudge's *intrusiveness* on Martin's behaviour. Participants were asked whether they found the intervention manipulative ('Do you think that the measure manipulates Martin's decisions?'), whether they thought the intervention restricted Martin's freedom of choice ('Do you think that the measure restricts Martin's freedom of choice?'), whether they thought Martin would approve of the interventions ('Do you think Martin would accept the intervention?'), whether they thought the intervention was ethical ('Do you think it is ethical to influence Martin's behaviour in this way?') and whether they thought the effect of the measure was positive for Martin ('Do you consider that the effect of the measure on Martin's behaviour is positive?'). All of the responses were rated on a continuous scale ranging from 0 to ('Not at all') to 100 ('Yes, totally').

Thirdly, a set of three questions collected respondents' *acceptability* of the nudges. Participants were asked whether they would support or oppose the measures with the following questions: 'Would you oppose the application of the measure?', 'Do you think the measure is acceptable?' and 'Would you be in favour of applying this measure?' Responses were rated on a continuous scale ranging from 0 to ('Not at all') to 100 ('Yes, totally').

Finally, the questionnaire also included a question on perceived effectiveness. ‘Do you think that the measure described will be effective?’ The question was rated on a continuous scale ranging from 0 to 100 (‘Not at all effective’) to 100 (‘Very effective’).

- *Manipulation checks*

Following the primary outcome measures, two questions were included to verify comprehension of the vignettes. The questions allowed me to assess whether participants' perceptions matched the conditions to which they were assigned. The *type of character* check verified whether participants understood the way in which Martin's character was described (‘How is Martin described?’ as: ‘Someone who is affected by context and does not mind letting go’ or ‘Someone who likes to be in control and make their own decisions’). The *type of influence* check verified whether participants understood the way in which the nudge's influence was described (‘How is the measure's influence on Martin's behaviour described?’ ‘Conscious’ or ‘unconscious’).

- *Co-variables questionnaire*

In addition to the central survey, the study included a post-experimental questionnaire on complementary predictors of acceptability. Participants answered the questionnaire after completing the survey. Firstly, the study collected *gender, age, education, activity* and *nationality* to consider the effect of sociodemographic characteristics. Secondly, the questionnaire collected *ideology* and *trust in public institutions* to control for the effect of political attitudes. Study B used the same questions as Study A for these factors. Thirdly, the questionnaire also asked individuals about their beliefs on their dispositions or situationism tendencies and those of others with the questions: ‘Do you consider that your behaviour is mainly determined by your choices and decisions or by the context and circumstances that surround you?’ and ‘Do you consider that the behaviour of most people is mainly determined by their choices and decisions or by the context and circumstances that surround them?’. The responses were ranked on a five-point Likert scale ranging from 1 (‘By my/their choices and decisions’) to 5 (‘By the context and circumstances that surround me/them’). Finally, two questions were included to collect people's beliefs and preferences about the policy goals considered in the vignettes. In the *healthy eating scenario*, participants were asked about their beliefs regarding the causes of obesity (‘Why do you think people are overweight and obese?’). The response was ranked on a five-point Likert scale ranging from 1 (‘Because there are too many opportunities to access unhealthy products and difficulties in leading a healthy life’) to 5 (‘Because people have self-control problems and do not correct their behaviour’). In the *organ donation scenario*, people were asked whether they were organ donors (‘Are you an organ donor?’, a question with two possible answers ‘yes’ or ‘no’). (See the questionnaire in Table 21, Appendix B).

- *Statistical analysis*

416 people conducted the experiment; 25 participants were excluded for not providing a coherent or readable answer to the open-ended question. Comprehension of the vignettes was verified by the responses to the questions checking manipulation. Several people failed to answer one or both of the manipulation checks correctly. Respondents with incorrect responses were excluded from the final analysis. The final sample consisted of 228 subjects. The sample had representative quotas in terms of gender (51% were female) and age (MD:40, sd:13.41). By contrast, the sample was better educated than the general Spanish population, with 62% of respondents with a university education. Likewise, people with full-time jobs were overrepresented: about 58% of the sample reported working full time. (See the characteristics of the participants in the study in Table 22, Appendix B).

The analysis used the responses to seven questions as the primary outcome measures. The internal consistency reliability of the nine original items was analysed using Cronbach's alpha. I took two scales into consideration: intrusiveness and acceptability. The intrusiveness scale consisted of six items: *general opinion about the vignette, manipulation, freedom of choice, type of effect, ethics, and perception of acceptance*. The scale was internally consistent, with all items worthy of retention ( $\alpha = .73$ ). The acceptability scale consisted of three items, support, acceptability and opposition to the intervention. The Cronbach's alpha revealed a high association between the items ( $\alpha = .98$ ), which indicates redundancy. Accordingly, I combined the responses to the three items in one single variable denoting *acceptability*. The reliability of the final seven items was strong ( $\alpha = .84$ ).

The seven response variables were analysed as continuous variables. One-way ANOVAs were used to explore the effect of the treatments. The data did not follow a normal distribution, but the sample for the different groups was sufficient to perform a parametric test ( $n >$  than 20) in almost all the cases. One-way ANOVAs were also used to explore the effect of the co-variables. All the analyses were coded and performed using R.

## **4.2. Results**

### *4.2.1. Treatment effects*

The One-way ANOVAs<sup>22</sup> reveal that the responses for many of the dependent variables change significantly between vignettes in both scenarios. However, overall, the differences are driven by the *type of nudge* and not by the *type of character*. In both the healthy eating and organ donation scenarios, the effect of the *type of nudge* is more relevant than the effect of the *type of character*.

---

<sup>22</sup>To ensure that the robustness of the analysis, I performed a non-parametric test for each of the response variables. The Kruskal-Wallis tests showed the same statistically significant difference between means as the one-way ANOVAs.

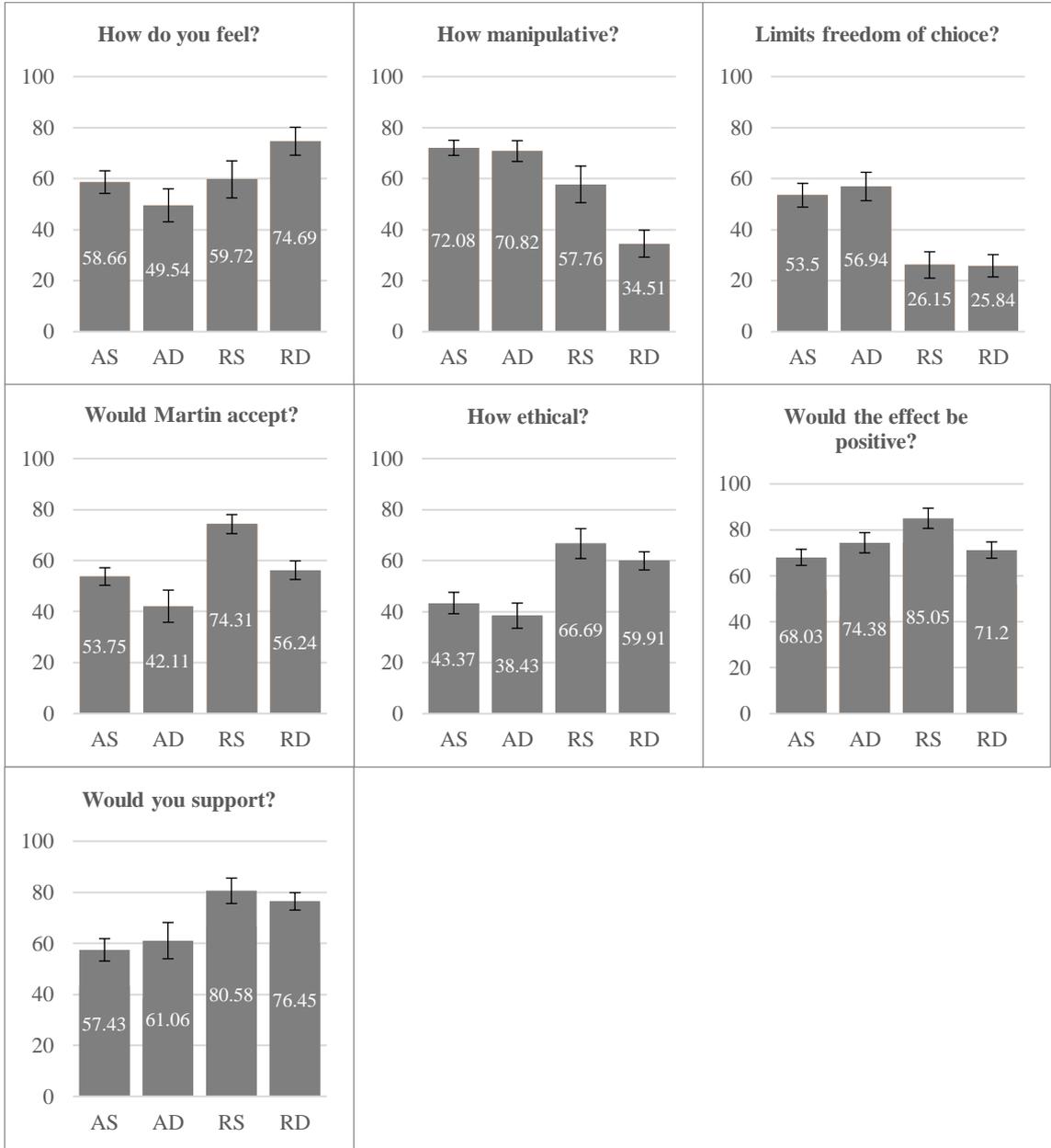
Figure 5 shows the means for each response variable for the healthy eating scenario. The one-way ANOVAs reveal a significant effect of the vignettes for every response variable. The descriptions in the vignettes affected how people felt about affecting Martin's behaviour as described in the vignette, ( $F(3, 107)= 2.919, p<0.05$ ) whether they perceived the influence as manipulative ( $F(3, 107)=15.14, p<0.01.$ ), whether they perceived the influence as limiting freedom of choice ( $F(3, 107)=9.962, p<0.01$ ), whether they thought Martin would accept the intervention ( $F(3, 107)=7.01, p<.01$ ), how ethical the situation was perceived ( $F(3, 107)=6.789, p<0.01.$ ), whether the effect was seen as positive or negative ( $F(3, 107)=2.986, p<0.00$ ) and whether they would support the use of the nudge ( $F(3, 107)=4.965, p<0.00$ ).

Pairwise comparisons of the means using Tukey's Honestly Significant Difference procedure indicated that differences between conditions were driven by *treatment 1. Type of nudge*. The AS-AD comparison was not significant in any case, suggesting that respondent's opinions on automatic nudges were unaffected by which type of decision they affected. Likewise, the comparison between RS-RD were minimal. They were only relevant for the questions on *manipulation* and *Martin's acceptability*. Subjects in the Reflective + Situationism condition reported that the influence was significantly more manipulative ( $p = .00$ ) than subjects in the Reflective + Dispositionist condition. By contrast, subjects in the Reflective + Situationist condition reported that Martin was significantly more likely to support the nudge ( $p = .04$ ) than subjects in the Reflective + Dispositions condition. (See the pairwise comparisons in Section 2.3, Appendix B).

Figure 6 shows the means for each response variable for the organ donation scenario. In terms of the treatment effects, results are very similar for scenario B, organ donation. The one-way ANOVA analysis reveal a significant effect of the vignettes for every response variable except for the question about how people feel about the situation described in the vignettes. The vignettes change whether people perceived the influence as manipulative ( $F(3, 113)=3.753, p<0.05$ ), whether people perceived the influence as limiting freedom of choice ( $F(3, 113)=4.76, p<0.001$ ), whether people thought Martin would accept the intervention ( $F(3, 113)=5.751, p<0.01$ ), how ethical the situation was perceived ( $F(3, 113)=5.931, p<0.001.$ ), whether the effect was seen as positive or negative ( $F(3, 113)=6.232, p<0.01$ ) and whether people would support the use of the nudge ( $F(3, 107)=3.547, p<0.05$ ).

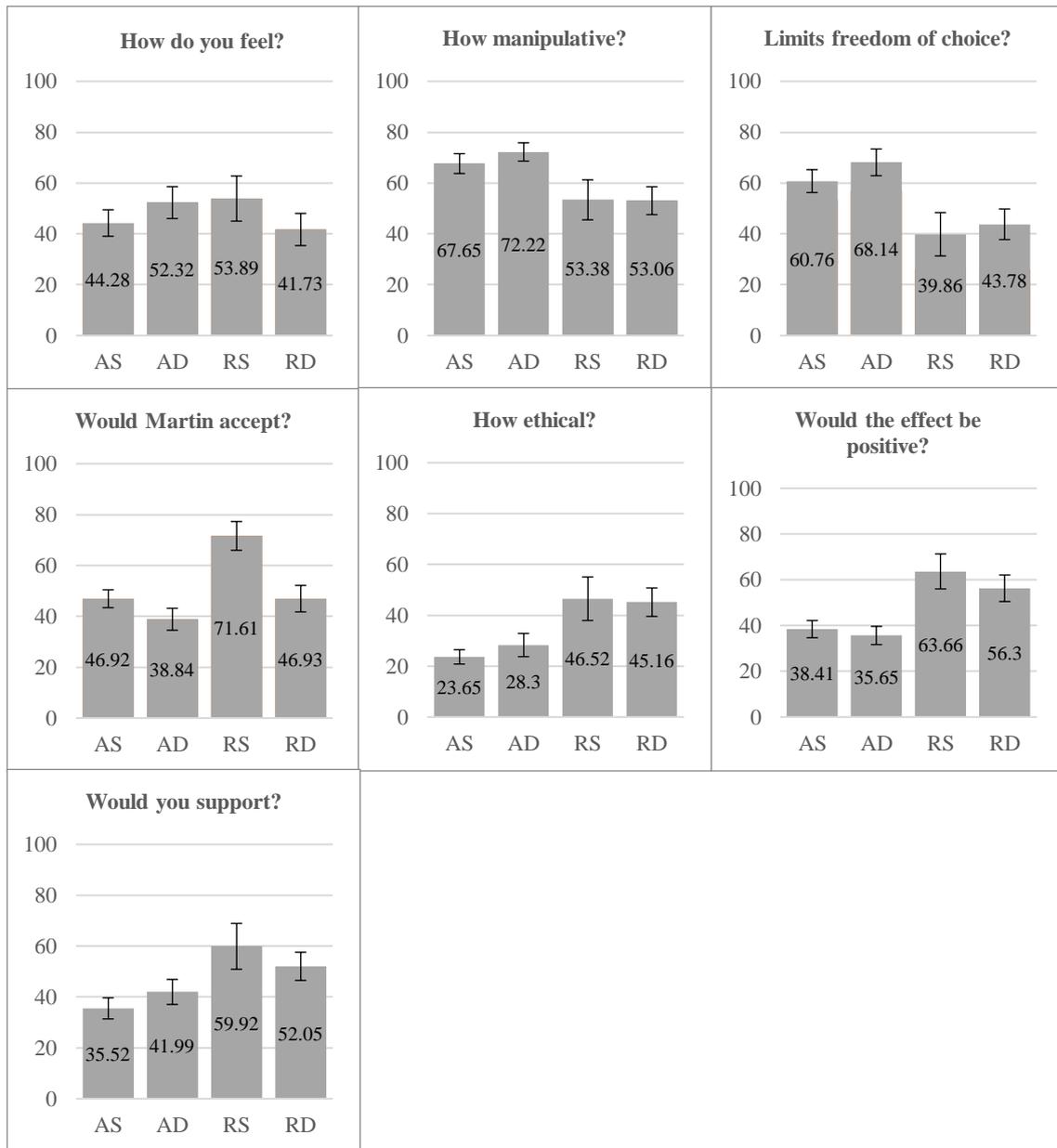
As in the healthy eating scenario, pairwise comparisons indicated that the differences between conditions were overall driven by *treatment 1. Type of nudge*. The AS-AD comparison was not significant in any case, suggesting, again, that respondents' perceptions about automatic nudges were unaffected by the type of decision they were making. The RS – RD comparison was only significant in the question: *Would Martin accept?* Subjects in the Reflective + Situationist condition were more likely to consider that Martin would accept the intervention ( $p=.01$ ) than subjects in the Reflective + Dispositionist condition. (See the pairwise comparisons in Section 2.3, Appendix B).

**Figure 5. Means for each response variable by condition in the healthy eating scenario.**



Scenario A. Healthy eating. Means and error bars for the seven response variables by experimental condition. x=experimental conditions, y=responses: 0=very bad, not manipulative, not limiting of freedom of choice, would not accept, very unethical, very negative, would not support. 100=very good, very manipulative, very limiting of freedom of choice, would accept, very ethical, very positive, would support.

**Figure 6. Means for each response variable by condition in the organ donation scenario.**



Scenario B. Organ donation. Means and error bars for the seven response variables by experimental condition. x=experimental conditions, y=responses: 0=very bad, not manipulative, not limiting of freedom of choice, would not accept, very unethical, very negative, would not support. 100=very good, very manipulative, very limiting of freedom of choice, would accept, very ethical, very positive, would support.

#### 4.2.2. Differences between scenarios

Overall, attitudes about the vignettes were more negative in the organ donation scenario than in the healthy eating scenario. People reported feeling worse about influencing Martin's behaviour to promote organ donation than healthy eating ( $t(226) = 3.26, p=0.0013$ ). Likewise, perception of whether the influence limited freedom of choice was more pronounced in the organ donation scenario than in the healthy eating scenario ( $t(226) = -3.07, p=0.0023$ ) and people also deemed Martin less likely to accept the nudge in the organ donation scenario than in the healthy eating scenario ( $t(226) = 2.33, p=0.02$ ). Differences were more noticeable in the *ethical*, *effect* and *support* questions. Influencing Martin's behaviour was perceived as being much more unethical in the organ donation scenario than in the healthy eating scenario ( $t(217) = 8.33, p<.00$ ) and less positive for Martin ( $t(225) = 4.86, p<.00$ ). As a result, people were less likely to support the nudges in the organ donation scenario than in the healthy eating scenario ( $t(226) = 5.806, p<.00$ ).

As shown in the last graphs (*Would you support?*) in Figures 5 and 6, the use of the nudges was always approved of in the healthy eating scenario, regardless of the type of nudge and type of character. By contrast, the use of nudges to promote organ donation was only approved of in the reflective conditions. Interestingly, the means of how manipulative the nudges were perceived were not statistically different between the two domains, although acceptability did vary. The results are consistent with hypothesis 5 and suggest that whether people would accept the use of automatic and reflective nudges does not only depend on the type of nudge but also on the aim the nudge is promoting. In the healthy eating scenario, people accepted the use of automatic nudges, even when they thought that their use is manipulative, that they limit freedom of choice, and that, overall, they are more unethical. Instead, people generally rejected the use of automatic nudges to promote organ donation.

#### 4.2.3. Co-variables effect

One-way ANOVAs were performed with co-variables to assess whether they had significant effects on opinions about the vignettes. The effect of sociodemographic factors, political values and individual disposition was minimal. No variable had a systematic effect except for *being an organ donor*, which had a systematic influence on subjects' responses to six out of the seven dependent variables in the organ donation scenario.

In the healthy eating scenario, minimal significant effects were found for gender, ideology and self-perception of dispositionism vs situationism. Women were more likely to feel good about influencing Martin's behaviour than men ( $p=.02$ ). The result is consistent with the findings that found women more in favour of intervention to promote healthy eating. Likewise, people who positioned themselves politically in the centre and on the right were significantly more likely ( $p=.04, p=.01$ ) to perceive that Martin would accept the intervention than those that positioned themselves on the left. Finally, people who felt strongly about their behaviour being mainly driven by their choices and dispositions were more

likely to judge the nudges as unethical than people who perceived themselves as somehow dispositionist but did not self-position themselves on the extreme (differences between 2 – 1  $p = .014$ ., differences between 3-1  $p = .008$ ).

In the organ donation scenario, minimal effects were found for education and self-perception of dispositions vs situationism. Those with a university education and those with a master's or PhD degree were more likely to value the effect of nudges as ethical than those with a secondary education (comparison between master's and PhD degrees – secondary education, ( $p = .06$ ) comparison between a university education and secondary education ( $p = .04$ )). Likewise, people who felt strongly about their behaviour being mostly driven by their choices and dispositions were more likely to feel good about influencing Martin's behaviour as described in the vignettes.

Being an organ donor had a systematic effect on people's opinions about nudges. Organ donors judged the interventions as less manipulative ( $F(1, 115) = 3.62, p < 0.01$ ) and less limiting on freedom of choice ( $F(1, 115) = 8.713, p < 0.01$ ). They were more likely to think that Martin would approve of the interventions ( $F(1, 115) = 3.862, p = 0.051$ ) and more likely to judge the effect on Martin's behaviour as positive ( $F(1, 115) = 3.62, p = 0.059$ ). Finally, they were more likely to perceive the interventions as ethical ( $F(1, 115) = 5.10, p = 0.02$ ) and more likely to support the interventions ( $F(1, 115) = 8.12, p < 0.01$ ).

## **5. Discussion**

Studies A and B investigated factors that affect the acceptability of automatic and reflective nudges. The results suggest that, in general, people always support the use of reflective nudges. By contrast, the acceptability of automatic nudges depends on the aim they are promoting and might vary according to how their effect is communicated. Study A and Study B show that people support the use of automatic nudges in the context of reducing the consumption of SSBs. By contrast, Study B shows that the majority of people do not support their use in the context of promoting organ donation. The results in Study B also indicate that negative attitudes towards automatic nudges are motivated by concerns about how they affect behaviour. However, these concerns do not have the same relevance when it comes to supporting or rejecting automatic nudges across domains.

Study A replicated previous research on the acceptability of policy measures to promote healthy choices. Results suggest that public acceptability of the measures appears to depend on the nature of the interventions. The study finds high support for educational and informative nudges, general acceptability for automatic nudges, but less approval for taxation.

Based on the findings of Study A, I found no evidence to suggest that highlighting the effectivity of automatic nudges increases their acceptability. Likewise, I found no evidence to suggest that highlighting the non-conscious nature of the mechanism by which automatic nudges work reduces their

acceptability. There are several possible explanations as to why the treatments were ineffective in changing the level of acceptability for the majority of interventions. The first explanation concerns problems with the design of the treatments. As commented, *Treatment 2. Time pressure* was problematic. It gave subjects too much time to evaluate the nudges and failed to induce a sensation of time pressure. As a result, there was no condition in which subjects had to assess the interventions under cognitive load. *Treatment 1. Information* might also have been ineffective due to problems with its design, probably related to the type of information, the wording and whether or not subjects understood and believed the information.

Previous experimental studies show that communicating evidence on effectiveness increases the acceptability of policy measures in several domains, including the acceptability of nudges used to change health-related behaviours. However, the effects are moderate and differ depending on how the evidence for effectiveness is communicated (Reynolds et al., 2020). Some studies indicate that simple information on effectiveness is sufficient. For instance, Reynolds et al. (2018) found that a simple message with unquantified assertions of effectiveness increased the acceptability of nudges and taxation to reduce the consumption of unhealthy food, tobacco and alcohol. The study also found that presenting the information in quantities did not make it more likely to improve acceptability. Sunstein (2017c) also found that communicating that System 1 nudges are “significantly more effective” than System 2 nudges increases the acceptability of System 1 nudges by 10% to 12%. Likewise, he also showed that adding quantified information had no further effect on acceptability. Davidai & Shafir (2018) found that in a joint evaluation, asserting the effectiveness of System 1 nudges as “significantly more effective” and presenting this difference with numeric information (40% vs 5% effectiveness) influenced participants’ acceptability of System 1 nudges. However, the same information had no significant influence on participants’ attitudes when nudges were evaluated separately.

Uncertainty remains about the most effective ways of presenting information about effectiveness. In Study A, the information indicated that interventions were effective at reducing the consumption of SSBs. The information was presented in qualitative terms and without expressing a magnitude (i.e., *strongly* or *somewhat* effective). It is possible that not indicating any magnitude made the wording too soft and, therefore, ineffective. Likewise, studies seem to suggest that the effect of communicating effectiveness on the acceptability of automatic nudges is more likely to be relevant when subjects have to compare interventions. In such cases, people tend to choose more effective interventions. In study A, interventions were evaluated separately, each on their own terms. Moreover, the information on effectiveness was the same for each intervention, and no measure was described as being more effective than any other, in order ensure consistency. As a result, a comparison between interventions was not favoured. This might explain why the information on effectiveness did not generally influence the acceptability of the measures.

Communicating effectiveness works as far as improving acceptability is concerned, because it updates beliefs about perceived effectiveness. The essential idea is that people tend to disapprove of interventions if they believe they are ineffective; therefore, when presented with evidence about their effectiveness, they update such beliefs and approve of the interventions (Reynolds et al., 2020). However, aside from the wording, there may be two reasons why information does not cause people to update their views. Firstly, when informed about the effectiveness of interventions, it is essential to trust the information and the source. Evers et al. (2018) found that the trustworthiness of the source implementing nudges to promote healthy choices affects public support for interventions and that that support increases when the nudges are implemented by scientists or private companies rather than public administrations. Study A listed the administration as the nudger, which might have triggered reservations about the credibility of the information.

Similarly, it is unclear how much people believed the information on effectiveness. It is not possible to know if the information on effectiveness modified respondents' perceived effectiveness because the conditions with information on effectiveness did not include a question on perceived effectiveness. In the open-ended questions, some respondents answered that they were assessing the acceptability of the measures with their subjective perception of how effective they would be, indicating that they did not believe the information they were provided. Therefore, it is possible that people did not believe the information and answered according to their perception of effectiveness.

Previous findings document a strong association between acceptability and perceived effectiveness. However, some authors have also questioned the relationship between the two factors. It is unclear whether perceived effectiveness precedes acceptability or if supporting an intervention makes people more likely to infer that it would be effective (Petrescu et al., 2016; Reynolds et al., 2018, 2020). As a result, motivated reasoning and confirmation bias might have played a role in whether people believe the information on effectiveness. These mechanisms might have made people more likely to ignore the evidence of effectiveness and perceived this issue according to how much they like or dislike the interventions<sup>23</sup>.

The problems with the effectiveness condition considered above help to explain why the information failed to improve the acceptability of nudges and subliminal messages. However, these considerations do not clarify why the treatment did work when it came to changing acceptability levels for the tax. The differences between the effect of the treatment between taxation and the other measures indicate that effectiveness may not be as important when considering their respective acceptability. Results on perceived effectiveness in conditions 0, 1, 4, and 5 suggest that people had unclear beliefs about the

---

<sup>23</sup>In cases in which people had clear and positive beliefs on the efficacy of the interventions (the case for educational and informational campaigns), they already favoured the interventions and considered the interventions to be effective. Thus, additional emphasis on the effectivity of the measures did not change acceptability levels.

efficacy of several of the policies. The percentage of neutral responses was around 30% for the measures related to size, location, app and subliminal messages. As a result, it is unclear how much weight people placed on effectiveness when it came to assessing the acceptability of these measures. By contrast, it is well documented that people believe that taxes are ineffective and that this issue is essential when they are considering whether they support or oppose their implementation. Results show that the tax was perceived as the least effective intervention, with 40% of respondents considering it ineffective (when 'strongly disagree' and 'disagree' responses<sup>24</sup> are combined). The fact that people had a more precise, negative perception of the effectiveness of the tax suggests that perceived effectiveness matters more when considering the acceptability of this measure. As a result, communicating that the tax was effective did change people's beliefs and had a positive influence on its acceptability.

Similar problems might explain why highlighting the nature of the mechanisms by which the interventions influence behaviour did not have an impact on the acceptability of measures related to nudges and subliminal messages. As for the case for effectiveness, the way of communicating the information and the way in which nudges were evaluated might have made the treatment ineffective.

Firstly, the words used to describe the non-conscious process might have been confusing or unsuccessful in terms of highlighting the implications of these interventions. Findings in other studies suggest that the wording matters. For instance, to compare preferences for covert and overt nudges, Felsen et al. (2013) describe the influence of covert nudges with the words "subconscious driven bias", terms with associated negative connotations. In their study, they found that people preferred System 2 nudges across different domains. In a similar study, Petrescu et al. (2016) analysed public attitudes towards System 1 and System 2. However, the authors use more neutral terms and employ the words "people will not be conscious (i.e., not aware)" to describe the influence of System 1 nudges. In their study, the authors found no effect of communicating this information and suggest that the toned-down description might have made the treatment ineffective. In Study A, the interventions related to size, location, default and subliminal messages were described with the words 'automatic and subconscious'. While it could be argued that both carry negative connotations, in Spanish, they can be confusing and might have been unclear at the time of communicating the implications of System 1 nudges. The wording, therefore, might have been ineffective in terms of influencing the acceptability of the different measures.

Similarly, as with the information on effectiveness, the fact that nudges were evaluated on their own and not compared to one another might have made the information on the mechanism less relevant. Davidai & Shafir (2018) provide evidence that joint evaluation accentuates preferences for System 2 nudges over System 1 nudges. However, in separate evaluations, the levels of support are similar. It is

---

<sup>24</sup>Responses in conditions 0, 1 (no information) and 4, 5 (information on the mechanism).

possible, therefore, that the influence of describing the mechanism of influence is only triggered in cases of joint evaluation and not when nudges are evaluated on their own, as in Study A.

The extent to which people perceived themselves as being affected by the interventions might also have been a problem. People have their own ideas on how interventions may affect their behaviour and the behaviour of others. Respondents may not have envisaged that the interventions would affect their behaviour subconsciously. 50% of respondents stated that their choices were mainly driven by their actions and disposition and not affected by the environment. By contrast, only 20% perceived themselves as someone whose choices were regularly affected by the context and circumstances around them. These results seem to suggest that respondents could be underestimating the capacity of some of the interventions to change their behaviour without their awareness. As a result, even when informed, some respondents might be unbothered by this information because they believe that the interventions will not affect them.

Finally, as extensively discussed throughout this thesis, the acceptability of nudges is mainly driven by the aims they promote and the context in which they are applied. In Study A, the majority of interventions were approved of. Likewise, highlighting the non-conscious nature of some measures did not decrease their acceptability. These results suggest that people agree with the implementation of these measures to reduce the consumption of SSBs, regardless of how they affect behaviour. Previous studies indicate that people approve of automatic nudges to promote healthy choices. While the levels of support are lower than for educational and informative interventions, people still support their use. In a very similar study examining preferences for intervention to reduce the consumption of SSBs, Petrescu et al. (2016) also found no effect of highlighting the non-conscious nature of the mechanisms. Their results are consistent with findings in Study A.

Two additional facts support this interpretation. Firstly, the implementation of *subliminal messages* had surprising levels of support. As commented, I expected more people to consider this intervention unacceptable. While surprising, previous studies have found minority support for subliminal messages for combating health risk behaviour (e.g. Sunstein, 2017c; Sunstein et al., 2018). These results suggest that when public support for the policy goal is high, some will even justify the use of strictly manipulative interventions. Secondly, in the open-ended questions, a majority of respondents stated worries about health and considerations about how to improve it as the most relevant factors they took into consideration when assessing the acceptability of the eight measures. These considerations support the idea that nudges are primarily judged based on their aim.

Additionally, I would like to point out that the survey-experiment started with a brief introduction that emphasised the individual and societal problems associated with being overweight and obesity. While the text was short and informative, it might have prompted respondents to judge the interventions in these terms and not according to their type of influence. Overall, the main finding is that even when

automatic nudges are described as interventions that affect behaviour automatically and subconsciously, people favour their implementation in the context of reducing the consumption of SSBs.

As with the information on effectiveness, information on the mechanism of influence did affect the acceptability of the tax. The mechanism condition described the influence of the tax as conscious, and the frame increased the acceptability of the measure. Two different factors may explain the positive impact of providing information about the mechanism. Firstly, as explained, the results in the control conditions suggest that people had negative attitudes towards the implementation of the tax. For this intervention, supporting the aim is insufficient when it comes to supporting its implementation. As a result, when respondents had more positive information about the measure, they were more likely to support it. Secondly, the intervention was perceived as one of the most manipulative. The emphasis on the fact that the tax works consciously may disperse worries about manipulation and, therefore, may have made people more likely to support the intervention. I did not expect concerns about manipulation to be so relevant for tax. While I have limited information to discuss these results, it could be the case that people also perceived the fact that taxes are often non-transparent or that people are not directly aware of their existence and effect. It appears that this is an issue that should be taken into consideration when assessing the acceptability of taxes. Finally, because people tend to prefer interventions that affect behaviour consciously, highlighting this aspect by itself might have been enough to improve the acceptability of the measures.

Study B confirms that people prefer reflective nudges but shows that the main factor in assessing the acceptability of automatic nudges is not how they affect behaviour but the aim they are promoting. The results indicate that, overall, automatic nudges are perceived as more intrusive than reflective nudges; however, perceptions about their intrusiveness and attitudes about their implementation are also shaped by the domain in which they are applied.

Study B provided evidence that attitudes towards automatic nudges are context dependent. Overall, automatic nudges were always perceived more negatively than reflective nudges, but with variations and different implications between domains. In the healthy eating scenario, the implementation of automatic nudges managed to gather the support of the majority. By contrast, in the organ donation scenario, the use of automatic nudges was disapproved of. The results indicate that people understand and acknowledge that automatic nudges are more intrusive on choices. However, the way in which automatic nudges affect behaviour is not always considered unacceptable. The results support hypothesis 5. Perceptions about nudges described in the same way are context dependent and are perceived as being more problematic in high-stake domains than in low-stake domains. The results are also consistent with the findings in Study A and reinforce the idea that people support the use of automatic nudges to promote healthy eating choices.

Perceptions of intrusiveness varied across the two domains. Automatic nudges were more negatively perceived in the organ donation scenario than in the healthy eating scenario. Primarily, their influence was seen as being more negative and perceived as being more limiting on freedom of choice, more unethical, less likely to be accepted by the nudgee and less favourable for whom is affected by the nudge. The results suggest that the evaluation of intrusiveness of automatic nudges is also affected by the domain in which they are applied. As argued in section 2, moral heuristics might influence people's judgments about automatic nudges. While I have limited information to confirm this hypothesis, the results in study B suggest that this might be the case. In contexts closer to sensitive issues or moral areas, such as organ donation, non-deliberative influences are judged more severely.

Contrary to the results in Study A, in Study B, *Treatment 1. Type of nudges* significantly influenced perceptions about automatic and reflective nudges. Highlighting that automatic nudges work through non-deliberative mechanisms and affect behaviour without the awareness of the nudgee increased their perception of intrusiveness and reduced their acceptability. The difference between the effect of similar treatments in Study A and Study B highlights the importance of the wording and the method of evaluation when assessing opinions about nudges. The vignettes may have made the treatment more effective because they allowed better and longer descriptions of the influence of the nudges and they described how they influenced a third person, which made the implications of automatic nudges more vivid.

*Treatment 2. Type of behaviour* was ineffective. As such, I found no evidence to support hypothesis 4. As in Study A, problems with its design might explain why it was ineffective. Firstly, the vignettes were somewhat unbalanced in terms of the concerns describing the type of nudge and character; there was more information on the nudge than on the type of character. As a result, the text might have made people focus more on the type of nudge than the type of behaviour they were influencing. Perhaps a more balanced description would have produced different results. Secondly, the scenarios in each vignette might have been somewhat difficult to understand. Several participants failed to answer the manipulation questions correctly, which indicates that comprehension of the vignettes was more difficult than anticipated. The errors followed a pattern. In general, it appears that people had a hard time understanding that someone described as a dispositionist could be influenced by a nudge, particularly by an automatic nudge, and they also had a hard time understanding that a reflective nudge would change the behaviour of a situationist. In general, the fact that the treatment was ineffective and that some subjects failed the control questions indicates that the vignettes were puzzling.

*Treatment 2. Type of behaviour* was intended to investigate whether automatic nudges would be accepted more readily when they influence unconscious and unintentional choices. In line with the philosophical arguments discussed in Chapter 2, the essential idea is that many of our choices tend to be unconscious or unintended and are already strongly influenced by our environments. In these contexts, influencing behaviour with automatic nudges appears to be less ethically problematic. Two

additional reasons might explain why the treatment failed to influence subjects' perceptions of the nudges. Firstly, the use of the *situationist vs dispositionist* tried to reflect the idea of hypothesis 4 by presenting a type of person who is often influenced by their environment in comparison with someone who is not. While the distinction partly captures the underlying intuition, it may have failed to reflect exactly what I intended. The treatment made the situation about the types of people we influence rather than about the types of choices we influence. Therefore, the treatment was ineffective when it came to emphasising the unconscious/unintended vs conscious/intended choices. Likewise, the wording used to describe *situationists* may have portrayed a somewhat naïve person. I can see how affecting the behaviour of a naïve person with automatic nudges can easily be perceived as problematic and manipulative. In future work, I intend to explore other ways of explaining how the environment influences choices and to see whether this factor shapes preferences for automatic nudges, or not.

Finally, perhaps my central intuition was wrong, and people do not judge automatic nudges as being less intrusive when they affect unconscious/unintended choices or when they are applied in contexts in which choice architecture heavily influences behaviour. However, I would like to remark that the three studies performed did produce a stable result, i.e., the notion that people do not perceive themselves as being affected by the context of choice. At the same time, they believe that the behaviour of most people is affected by it. It appears that people do acknowledge the relevance of the context of choice in shaping choices, yet they do not recognise or do not like to believe that it also affects them. In this context, studying how the perception of the environment shapes the acceptability of nudges becomes more challenging, because people may not believe or perceive the influences as affecting them. A possible way of resolving this issue is to reformulate the vignettes, so that they further emphasise the effect of the environment, and to present a second person rather than a third person view of the effect. Overall, I found no evidence to support hypothesis 4, but these results suggest that further research is needed to reject it.

As far as independent predictors of acceptability are concerned, Study A and Study B found no systematic effect of any variable. However, a relevant pattern emerges. Individual preferences related to the aims of the nudge affect their acceptability. As regards the acceptability of nudges in the healthy eating scenario, both studies showed that women are more likely to support interventions to promote healthy eating choices, probably because they are more concerned with this issue. Likewise, people that reported being on a diet were also more likely to support some of the nudges to reduce the consumption of SSBs. By contrast, those that self-reported regular consumption of SSBs were less likely to support the measures. Similarly, in the organ donation scenario, being an organ donor systematically affected people's opinions about the nudges, with donors being more in favour of their implementation. These results concur with previous evidence and underline the importance of the relationship between supporting the aim of the nudge and accepting their use.

Recent debate among proponents and critics of nudging has centred on the ethical legitimacy of using automatic nudges. The philosophical debate has been searching for ways to classify nudges according to the mechanisms by means of which they work and judging whether they should or should not be used, depending on different normative understandings of decision-making. Recent literature on attitudes towards nudges has begun to study public support for their implementation, contrasting their use with the implementation of more deliberative interventions. The literature on attitudes towards nudges has underlined the importance of considering people's opinions when it comes to resolving the relevant ethical questions surrounding the permissibility of automatic nudges. However, the questions are still too centred on the way in which automatic nudges affect behaviour. By contrast, I suggest that the debate should not only focus on assessing attitudes towards nudging solely based on how they work, but also on exploring the contexts in which influencing choices with automatic nudges is considered permissible or impermissible.

By studying attitudes towards automatic nudges in comparison to reflective nudges, previous research might have overstated negative opinions about automatic nudges. We have evidence that people prefer reflective nudges to automatic nudges. However, people also prefer educational and informational measures for regulation, taxes and other more intrusive policies. In general, reflective nudges are preferred, but this does not imply that automatic nudges are disapproved of. The main strength of the two studies is that they present evidence that shows that opinions about automatic nudges are affected not only by the way in which they work, but also by the domain in which they are applied, and the type of choices they affect and promote.

The findings provide three relevant insights for the debate on the acceptability of automatic nudges and assessing their policy implications. Firstly, they provide evidence that the use of automatic nudges is accepted in the context of reducing the consumption of sugar-sweetened beverages. They also indicate that highlighting that these measures work through non-conscious routes does not reduce support for their implementation to promote this aim. Secondly, the studies indicate that automatic nudges are not considered unacceptable per se, it is rather that their permissibility depends on the domain in which they are applied. They are accepted when it comes to promoting healthy choices, but unaccepted when used to promote organ donation. I suggest that the difference between domains is caused by the fact that healthy eating is a low-stakes domain. We make many choices about what to eat or drink, they are easily reversible, and they tend not to involve moral, religious or political values. By contrast, organ donation is a high-stakes domain, a decision that we probably only take once, and one that involves moral and religious values. As a result, we care more about making deliberative choices regarding organ donation. The pro-self vs pro-social distinction may also explain the differences between domains. Because healthy eating promotes individual wellbeing and organ donations does not, the willingness to accept automatic nudges may be lower. Finally, perceptions about the intrusiveness of automatic nudges vary between domains. Influences described in the same way are more negatively perceived in the high-

stakes domain than in the low-stakes domain. These results might suggest that our evaluations on whether to accept automatic nudges or not might be affected by System 1 and by moral heuristics. While I have limited evidence to support this claim, I suggest that it is a factor that should be taken into consideration in future studies.

It is worth noting some of the limitations of these studies. Firstly, an essential weakness of Study A and Study B is that they have somewhat small samples. As regards Study A, previous research has found that information on effectiveness and on the mechanisms by means of which nudges affect behaviour tends to have a small or moderate effect on public support for nudges (e.g. Davidai & Shafir, 2018; Reynolds et al., 2020; Sunstein, 2017c). Therefore, it is possible that the sample may be too small to detect changes on the acceptability of such an effect. As regards Study B, some subjects were excluded from the final analysis and the sample for some groups was small due to problems of understanding. Again, with a larger sample, results may differ. Another problem in both studies is that the questions and the vignettes assess attitudes about hypothetical policies and scenarios. The possibility remains that respondents' perceptions of nudges may differ from how they would actually feel in a real situation.

Despite these limitations, the studies provide evidence that attitudes towards automatic nudges vary across domains and contribute to emphasising the importance of including people's opinions in the discussion about nudge acceptability. The studies also provide insights that need to be explored in subsequent studies. A further focus for future studies is to consider how current findings generalise across other contexts and to study how people feel about implementing automatic nudges in high-stakes and low-stakes domains and use them to promote pro-self and pro-social goals. Likewise, the studies suggest that the way in which people understand the influence of their environment on choices is a factor that should be taken into consideration. Further studies exploring this issue may contribute to learning how to frame policy interventions to maximise public support for them.

## **6. Concluding remarks**

Chapter 7 explores factors that affect the acceptability of automatic nudges and reflective nudges. It provides relevant insights for policymaking by showing that people accept the use of automatic nudges in the context of reducing the consumption of sugar-sweetened beverages. The results suggest that when nudges are used to promote a supported outcome and in a relatively low-stakes context, automatic nudges might garner public support. Likewise, the findings indicate that people do not judge automatic nudges solely by the mechanisms by which they affect behaviour. However, they take the aims they are promoting and the domains in which they are applied into consideration. This factor is crucial when it comes to considering where and how to use automatic nudges in policymaking.

# Conclusions

## 1. Main contributions

The acceptability of nudges as policy tools is an issue that involves considerations on behaviour and decision-making, and considerations regarding how policy instruments should work, and which aims they should promote. Their implications in terms of policymaking are positive overall, but specific issues have to be taken into consideration. This thesis discusses a series of elements that are important when it comes to understanding how nudges should be applied in policymaking to ensure their acceptability.

- 1) *Rejecting the original nudge characterisation.* The first main conclusion of the thesis is that the justification and legitimisation of nudges in policymaking need not be based on the aims and values established in their original definition. In *'Nudge'*, Thaler and Sunstein define nudges as a policy instrument to change people's choices, but also characterise nudges as interventions that promote subjective wellbeing and respect freedom of choice. The association of nudges with specific values and aims is, however, problematic. Instead of providing a strong legitimisation of nudges, the original approach gives rise to significant criticisms and limits the conversation on how nudges may contribute to policymaking beyond their original intentions. As detailed throughout the thesis, the welfare-promoting and freedom-preserving properties of nudges cannot be generally maintained without raising problematic issues. Overall, the main objections to the acceptability of nudges suggest that their original normative intention is doomed and, therefore, should be rejected.
- 2) *Defining nudges according to their features as behavioural change interventions.* Rejecting the original characterisation of nudges leads to conceptualisation problems. Because nudges were initially conceived as interventions that promote subjects' wellbeing and respect freedom of choice, when these two properties are excluded, it becomes unclear what a nudge is exactly. In response, the thesis proposes a definition of nudges that trims down the normative content and defines nudges only according to the features that are essential to understanding how they affect people's choices. The thesis argues that what is novel and interesting about nudges is not the type of normative goals they are expected to promote, or whether they can promote subjective wellbeing and respect freedom of choice. Instead, what is important for policymaking is the idea that we can change behaviour by appealing to cognitive and social factors. By addressing the lack of clarity over what the term nudge is supposed to refer to, the thesis proposes defining nudges according to three main properties. A nudge: (i) does not change economic incentives, (ii) does not forbid any options, and (iii) affects the behaviour of agents by appealing to factors related to biases and blunders, temptation and/or following the herd. The definition is based on the original understanding of nudges, but it explicitly references the mechanisms that nudges are expected to tap into to affect people's choices.

The definition focuses on the practical value of nudges, which is a better starting point when it comes to discussing their potential positive or negative implications.

- 3) *Nudges are not intrinsically problematic but require discussion regarding how and where they should be applied.* Overall, the thesis suggests that nudges are not intrinsically problematic in any way. However, it does not attempt to argue that their implementation is always ethically permissible or that nudges are a better form of policy interference than other policy measures. Nudges are a tool with distinctive features, and, in some cases, their implementation might have negative impacts. As with any other policy tool, the potential negative implications of nudges should be considered.
- 4) *The acceptability of nudges is an issue that requires insights on how people feel about their implementation.* The debate on the acceptability of nudges is dominated by abstract notions and philosophical ideas that tend to express conflicted views and, as a result, is inconclusive. The thesis suggests that a good way to complement the theoretical and philosophical discussion and identify the way in which nudges should be implemented is to consider how people feel about the use of nudges.
- 5) *The implementation of nudges is generally supported in Spain.* The empirical results of the thesis are positive for the implementation of nudges in Spain. The survey carried out as part of the thesis reveals that acceptability of nudges is high; people favour the use of different nudges applied to both pro-self and pro-social domains and indicates that people do not categorically oppose nudges, do not perceive nudges as intrinsically manipulative and value them as a positive tool for behavioural change. The results resemble the findings in other European countries and western democracies and suggest that Spain belongs in the “principled pro-nudge nations” category, with general support for nudge interventions, but opposition to some interventions. Likewise, the findings suggest that all types of citizens may support nudges. The results indicate that sociodemographic characteristics and political views do not seem to be a significant determinant in the public’s views on nudges.
- 6) *Evaluating the acceptability of a nudge requires considering the aim of the nudge, the type of nudge and the context in which the nudge is applied.* The thesis argues and provides empirical evidence that the acceptability of nudges in policymaking depends on what nudges are used for, which nudge is being implemented and in which context. These three factors are all important and none of them is enough by itself to provide conclusive answers on the acceptability of nudges.
  - a. *The aim of nudges is a crucial factor in ensuring their acceptability.* What nudges are used for is one of the most important elements when it comes to assessing and ensuring their acceptability. Like any other policy tools, nudges should be used to encourage legitimate aims, and their implementation should be subject to standard democratic discussion and control. Findings on public attitudes towards nudges indicate that a nudge’s aim is the most relevant

factor when considering whether it is acceptable or unacceptable. People tend to support nudges when they agree with their aims and tend to reject them when they disagree with them. As a result, ensuring that people support the aim is crucial for the acceptability of nudges.

- b. *The acceptability of pro-self and pro-social nudges follows different standards.* The results of the studies conducted in the thesis indicate that pro-self nudges tend to be more supported, whereas pro-social nudges have more modest approval rates. However, the results also show that adverse reactions are stronger for pro-self nudges. This finding, while merely preliminary, suggests that support for pro-self and pro-social nudges follows different standards. The acceptability of pro-social nudges might be related to how much people care or agree with a specific policy goal. The lower acceptability rates are to be expected because pro-social nudges tend to be applied in areas of public policy that have a social dilemma structure, and in which common good requires some sacrifice or cost for individual wellbeing. In line with these results, the thesis points out that it may be appropriate to implement nudges to promote accepted and legitimised pro-social goals or to improve the performance of policies and programmes already implemented to reduce legitimisation problems. By contrast, pro-self nudges are generally supported but more likely to be rejected when people disagree with their aim. Along these lines, I suggest that people may favour pro-self nudges, not because they accept their use in changing their behaviour, but because they perceive themselves as not particularly bothered by them. As a result, it is crucial to know how nudges feel about implementing pro-self nudges in order to implement them and ensure their acceptability. The findings indicate that it is particularly important to consider people's opinions when implementing pro-self nudges and identify the contexts in which they require external support and, therefore, will welcome a pro-self nudge.
- c. *Nudges cannot be evaluated as a homogeneous category.* As argued throughout the discussion, one of the main worries about nudges is their potential negative impact on individual autonomy due to the mechanisms through which they affect behaviour. In response to this claim, the thesis emphasises that nudges cannot be evaluated as a homogenous category. Nudges operate in different ways, and considering these differences is crucial when it comes to assessing their ethical acceptability. The thesis proposes classifying nudges into two categories: automatic and reflective nudges. Automatic nudges are less noticeable, work with no deliberation and awareness, and have to be attached to the choice execution context to be effective. By contrast, reflective nudges are noticeable, work by engaging deliberation and do not have to be attached to the choice execution context to work. Automatic nudges are not ethically problematic *per se*, but they are more likely to affect agents' behaviour without them being aware of the nudge. The empirical findings of the three studies support the importance of taking the type of nudge into account in order to assess its ethical implications. People recognise differences between nudges and tend to prefer reflective nudges to automatic nudges. At the same time, the findings

of the three studies also indicate that people do not categorically oppose the implementation of automatic nudges and, in several cases, they actually support their use. In some cases, the results suggest that people agree with the implementation of automatic nudges when they favour the aim. Likewise, the findings indicate that the acceptability of nudges is affected by the context in which they are applied.

- d. *The acceptability of automatic nudges should not be evaluated by comparing it to the acceptability of reflective nudges.* The literature on attitudes towards nudges has underlined the importance of considering people's opinions in order to resolve the relevant ethical questions around the permissibility of automatic nudges. However, the questions are still too centred on how automatic nudges affect behaviour and studies consider their acceptability in comparison to the implementation of more deliberative interventions. By contrast, I suggest that the debate should not only focus on assessing attitudes towards nudging according to how they work but should also explore the contexts in which it is considered permissible or impermissible to influence choices using automatic nudges. Previous research might have overstated negative opinions about automatic nudges when studying attitudes towards automatic nudges in comparison to reflective nudges. In general, reflective nudges are preferred, but this does not imply that automatic nudges are disapproved of. One of the main strengths of the studies performed in the thesis is that they present evidence that shows that opinions about automatic nudges are affected by how they work, but also by the domain in which they are applied and the type of choices that they affect and promote. In some cases, the implementation of automatic nudges is seen as unproblematic.
- e. *The ethical acceptability of automatic nudges should be evaluated by taking the context in which an automatic nudge is applied into consideration.* The implications of nudges, particularly automatic ones, depend on the context in which they are applied. Research indicates that choices are heavily shaped by the context of choice and by the influence of intentional and unintentional factors that are often not directly evident to agents. The objections to nudges that are concerned with how they work tend to put too much emphasis on how nudges affect reflection and decision-making and tend to overlook how retaining process autonomy might be counterproductive to outcome autonomy and agents' overall welfare. Along the same lines, this thesis argues that influencing behaviour through automatic nudges is less problematic when these nudges affect unconscious/unintended choices and are applied in contexts in which choice architecture heavily influences behaviour. Likewise, the thesis argues that making choices and decisions is important, but it is not all equally relevant across domains. Along similar lines, it is argued that the acceptability of automatic nudges is less problematic when they are used to change behaviour in low-stake domains and more ethically problematic when they are used in high-stake domains. The empirical findings of the three studies partly support these arguments.

Firstly, they provide evidence that the use of automatic nudges is accepted in the contexts in which people support the aim, or in contexts in which people might have perceived that others have self-control problems. Secondly, the findings indicate that perceptions about the intrusiveness of automatic nudges vary between domains. The results of Study B suggest that influences described in the same way are more negatively perceived in high-stake domains than in low-stake domains, recognising the relevance of the context in which the nudges are applied. Study A indicates that highlighting the fact that these measures work through non-conscious routes does not reduce support for their implementation. So, overall, the studies indicate that automatic nudges are not considered unacceptable per se, rather their permissibility depends on the domain in which they are applied.

## **2. Limitations and possibilities for future research**

This thesis examines the acceptability of automatic nudges with limited consideration of some factors that are also important for this issue. One of the issues that the thesis has not considered, but that is relevant when it comes to assessing whether nudges are good policy measures is their long-term implications. I decided not to include a discussion on this issue in the thesis to avoid speculation, because this is fundamentally an empirical question and evidence on the long-term effects of nudges is not available. However, how nudges affect long-term decision-making, whether their effects persist and how they may affect agents' decision-making are all elements that are relevant when it comes to considering whether nudges are adequate policy tools. The implementation of nudges is relatively new, and we still lack a proper understanding of their potential long-term implications. This is an issue that needs to be empirically tested and included in future discussions about the acceptability of nudges.

The thesis contends that the acceptability of nudges and people's attitudes towards nudges are affected by the aim of a nudge and the context in which the nudge is applied. It also provides considerations on how people feel about nudges in specific domains, particularly the healthy eating and organ donation domains. However, many aims, domains and contexts are unexplored. A further focus for future studies could be to consider how current findings generalise to other contexts and study how people feel about implementing nudges in high-stake and low-stake domains and using nudges to promote pro-self and pro-social goals. In the case of automatic nudges, exploring their public support in different contexts is crucial, as their acceptability is particularly sensitive to the context in which they are used, and what they are promoting.

The thesis also has some methodological limitations that should be noted. Firstly, the three empirical studies assess and report attitudes about hypothetical policies and scenarios. The possibility remains that respondents' perceptions of nudges may differ from how they would actually feel when a nudge is applied in a real-life context, and when nudges affect their real choices. This is an issue that needs further testing. It would be interesting for field experiments testing the effectiveness of nudges to also

include some measurement of or questions about how people feel about being nudged. This would provide relevant insights for the discussion on the acceptability of nudges and make it possible to see whether the trends identified in the attitudes towards nudges literature match people's real feelings about nudges. Another weakness of Study A and Study B is that they have somewhat small samples. Regarding Study A, previous research has found that information on effectiveness and on the mechanisms through which nudges affect behaviour tends to lead to small or moderate effects on public support for nudges. Therefore, it is possible that the sample was too small to detect changes in acceptability. Regarding Study B, some subjects were excluded from the final analysis due to problems of understanding, and the sample for some groups was small. Again, results might differ with a larger sample. Overall, it would be beneficial to conduct more studies on how information on effectiveness and the influence of the context of choice affects the acceptability of nudges in order to complement the findings of Study A and Study B.

Finally, considering the ideas in the philosophical discussion, the thesis explores whether moral heuristics affect people's opinions about nudges and whether people's opinions about automatic nudges change when they affect unconscious/unintended choice. Overall, the studies found no evidence to support these ideas, but these issues require further testing. A stable result in the three studies performed demonstrated the notion that people do not perceive themselves to be affected by the context of choice. At the same time, they believe that the behaviour of most other people is. It appears that people do acknowledge the relevance of the context of choice in shaping choices, yet they do not recognise or do not like to believe that it also affects themselves. In this context, studying how people's perception of the environment shapes the acceptability of nudges becomes a crucial element. Likewise, updating their beliefs about this issue may contribute to knowing how to frame policy interventions to maximise their public support. Overall, further empirical testing of these two issues appears important when it comes to understanding attitudes towards nudges and other policy measures.

### **3. Final thought**

Discussing the acceptability of nudges is important when it comes to identifying how they should be used in policymaking. This work contributes to this end. Firstly, the thesis helps to refute the original criticisms and emphasise the practical value of nudges in policymaking by looking at the problems related to the original understanding of nudges and providing a new definition of the concept. Secondly, the thesis is able to discuss potential problems and how to resolve them from this new understanding, with particular emphasis on the importance of considering how people feel about the use of nudges in order to discuss their acceptability. The thesis also contributes to the literature on attitudes towards nudges and provides relevant insights in order to advance this research. One can only hope that it may contribute towards 'nudging' public agencies and administrations into paying more attention to the potential of using the tool to achieve valuable policy aims.

## References

- Afif, Z., Islan, W. W., Calvo-Gonzalez, O., & Dalton, A. G. (2018). *Behavioral science around the world: Profiles of 10 countries*. The World Bank.
- 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>
- Angner, E. (2015). To navigate safely in the vast sea of empirical facts. *Synthese*, 192(11), 3557–3575.
- Angner, E. (2016). *A Course in Behavioral Economics*.
- Angner, E., & Loewenstein, G. (2007). Behavioral economics. *Handbook of the Philosophy of Science: Philosophy of Economic*, 641–690.
- Arad, A., & Rubinstein, A. (2018). The People’s Perspective on Libertarian-Paternalistic Policies. *The Journal of Law and Economics*, 61(2), 311–333. <https://doi.org/10.1086/698608>
- Baldwin, R. (2014). From Regulation to Behaviour Change: Giving Nudge the Third Degree. *The Modern Law Review*, 77(6), 831–857.
- Behavioural Insights Team. (2019). *Annual Report 2017–18*.
- Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should Governments Invest More in Nudging? *Psychological Science*, 28(8), 1041–1055. <https://doi.org/10.1177/0956797617702501>
- Bhargava, S., & Loewenstein, G. (2015). Behavioral economics and public policy 102: Beyond nudging. *American Economic Review*, 105(5), 396–401.
- Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
- Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.
- Bleich, S. N., Herring, B. J., Flagg, D. D., & Gary-Webb, T. L. (2012). Reduction in purchases of sugar-sweetened beverages among low-income black adolescents after exposure to caloric information. *American Journal of Public Health*, 102(2), 329–335. <https://doi.org/10.2105/AJPH.2011.300350>
- Blumenthal-Barby, J. S. (2013). Choice Architecture: Improving Choice While Preserving Liberty? In C. Coons & M. Weber (Eds.), *Paternalism*. Cambridge University Press.
- Blumenthal-Barby, J. S., & Burroughs, H. (2012). Seeking better health care outcomes: The ethics of using the “nudge.” *American Journal of Bioethics*, 12(2), 1–10. <https://doi.org/10.1080/15265161.2011.634481>
- Bobadilla-Suarez, S., Sunstein, C. R., & Sharot, T. (2017). The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses. *Journal of risk and uncertainty*, 54(3), 187–202.
- Borgmeier, I., & Westenhoefer, J. (2009). Impact of different food label formats on healthiness evaluation and food choice of consumers: A randomized-controlled study. *BMC Public Health*, 9, 184. <https://doi.org/10.1186/1471-2458-9-184>

- Bovens, L. (2009). The Ethics of Nudge. In T. Grüne-Yanoff & S. O. ; Hansson (Eds.), *Preference change. Approaches from philosophy, economics and psychology* (Springer, pp. 207–219). Springer. [https://doi.org/10.1007/978-90-481-2593-7\\_10](https://doi.org/10.1007/978-90-481-2593-7_10)
- Bovens, L. (2013). Why couldn't I be nudged to dislike a Big Mac? *Journal of Medical Ethics*, 39(8), 495–496. <https://doi.org/10.1136/medethics-2012-101110>
- Bradt, J. (2019). Comparing the effects of behaviorally informed interventions on flood insurance demand: An experimental analysis of 'boosts' and 'nudges.' *Behavioural Public Policy*, 1–31.
- Briggs, R. A. (2019). Normative Theories of Rational Choice: Expected Utility. In *The Stanford Encyclopedia of Philosophy: Vol. Fall 2019 Edition* (Edward N. Zalta). Metaphysics Research Lab, Stanford University. <<https://plato.stanford.edu/archives/fall2019/entries/rationality-normative-utility/>>
- Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Luistro, M., & Rahali, B. (2018). Can nudges be transparent and yet effective? *Journal of Economic Psychology*, 65(February), 41–59. <https://doi.org/10.1016/j.joep.2018.02.002>
- Capraro, V., Everett, J. A., & Earp, B. D. (2019). Priming intuition decreases endorsement of instrumental harm but not impartial beneficence. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3315385>
- Chapman, B. A. (2016). The Ethics of Nudge: From Liberty to Autonomy. *ProQuest Dissertations and Theses*, 388.
- Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5), 1–33.
- Clery, E., & Rhead, R. (2013). Education and attitudes towards the environment. *Background Paper Prepared for the Education for All Global Monitoring Report, 4*, 2013.
- Conly, S. (2013). *Against autonomy: Justifying coercive paternalism*. Cambridge University Press.
- Conly, S. (2016). When freedom of choice doesn't matter. *The Tocqueville Review/La Revue Tocqueville*, 37(1), 39–58.
- Curchin, K. (2017). Using Behavioural Insights to Argue for a Stronger Social Safety Net: Beyond Libertarian Paternalism. *Journal of Social Policy*, 46(2), 231–249. <https://doi.org/10.1017/S0047279416000672>
- Datta, S., & Mullainathan, S. (2014). Behavioral Design: A New Approach to Development Policy. *Review of Income and Wealth*, 60(1), 7–35. <https://doi.org/10.1111/roiw.12093>
- Davidai, S., & Shafir, E. (2018). Are 'nudges' getting a fair shot? Joint versus separate evaluation. *Behavioural Public Policy*, 1–19. <https://doi.org/10.1017/bpp.2018.9>
- Dayan, E., & Bar-Hillel, M. (2011). Nudge to nobesity II: Menu positions influence food orders. *Judgment and Decision Making*, 6(4), 333–342.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2), 315–372.
- Dhami, S., al-Nowaihi, A., & Sunstein, C. R. (2019). Heuristics and public policy: Decision-making under bounded rationality. *Studies in Microeconomics*, 7(1), 7–58.
- Dworkin, G. (1988). *The theory and practice of autonomy*. Cambridge University Press.

- Elqayam, S., & Evans, J. S. B. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, *34*(5), 233.
- Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge University Press.
- Elster, J. (2000). *Ulysses unbound: Studies in rationality, precommitment, and constraints*. Cambridge University Press.
- Elster, J. (2007). *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*.
- Engelen, B. (2019). Nudging and rationality: What is there to worry? *Rationality and Society*, *31*(2), 205–232.
- Epley, N., & Tannenbaum, D. (2017). Treating ethics as a design problem. *Behavioral Science & Policy*, *3*(2), 72-84.
- Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, *7*(10), 454-459.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, *59*, 255-278.
- Evers, C., Marchiori, D. R., Junghans, A. F., Cremers, J., & De Ridder, D. T. D. (2018). Citizen approval of nudging interventions promoting healthy eating: The role of intrusiveness and trustworthiness. *BMC Public Health*, *18*(1), 1–10. <https://doi.org/10.1186/s12889-018-6097-y>
- Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: Public attitudes towards overt and covert nudges. *Judgment and Decision Making*, *8*(3).
- Felsen, G., & Reiner, P. B. (2011). How the neuroscience of decision making informs our conception of autonomy. *AJOB Neuroscience*, *2*(3), 3–14. <https://doi.org/10.1080/21507740.2011.580489>
- Felsen, G., & Reiner, P. B. (2015). What can Neuroscience Contribute to the Debate Over Nudging? *Review of Philosophy and Psychology*, *6*(3), 469–479. <https://doi.org/10.1007/s13164-015-0240-9>
- Fischer, M., & Lotz, S. (2014). Is Soft Paternalism Ethically Legitimate? - The Relevance of Psychological Processes for the Assessment of Nudge-Based Policies. *Cologne Graduate School Working Paper Series*, *5*(2), 17.
- Franklin, M., Folke, T., & Ruggeri, K. (2019). Optimising nudges and boosts for financial decisions under uncertainty. *Palgrave Communications*, *5*(1), 1–13.
- Gabaix, X., & Laibson, D. (2017). *Myopia and discounting* (No. 0898–2937). National bureau of economic research.
- Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain? *Journal of Consumer Psychology*, *28*(3), 497–516.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases.” *European Review of Social Psychology*, *2*(1), 83–115.
- Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. Oxford University Press.
- Gigerenzer, G. (2015). On the Supposed Evidence for Libertarian Paternalism. *Review of Philosophy and Psychology*, *6*(3), 361–383. <https://doi.org/10.1007/s13164-015-0248-1>

- Gigerenzer, G. (2018). The bias bias in behavioral economics. *Review of Behavioral Economics*, 5(3–4), 303–336.
- Gigerenzer, G., & Todd, P. M. (1999). Simple heuristics that make us smart. Oxford University Press, USA.
- Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 25(4), 191–210. <https://doi.org/10.1257/jep.25.4.191>
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107(3), 1144–1154.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108.
- Grüne-Yanoff, T. (2016). Why Behavioural Policy Needs Mechanistic Evidence. *Economics and Philosophy*, 32(03), 463–483. <https://doi.org/10.1017/S0266267115000425>
- Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge Versus Boost: How Coherent are Policy and Theory? *Minds and Machines*, 26(1–2), 149–183. <https://doi.org/10.1007/s11023-015-9367-9>
- Grüne-Yanoff, T., Marchionni, C., & Feufel, M. A. (2018). *Toward a framework for selecting behavioural policies: How to choose between boosts and nudges*.
- Guala, F. (2000). The logic of normative falsification: Rationality and experiments in decision theory. *Journal of Economic Methodology*, 7(1), 59–93.
- 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>
- Gyrd-Hansen, D., & Kjær, T. (2015). Government interventions to aid choice: Help to self-help or paternalism? *Health Policy*, 119(7), 874–881. <https://doi.org/10.1016/j.healthpol.2015.02.001>
- Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public Views on Policies Involving Nudges. *Review of Philosophy and Psychology*, 6(3), 439–453. <https://doi.org/10.1007/s13164-015-0263-2>
- Hagmann, D., Siegrist, M., & Hartmann, C. (2018). Taxes, labels, or nudges? Public acceptance of various interventions designed to reduce sugar intake. *Food Policy*, 79, 156–165. <https://doi.org/10.1016/j.foodpol.2018.06.008>
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814.
- 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.
- Hausman, D. M. (2018). Nudging and Other Ways of Steering Choices. *Intereconomics*, 53(1), 17–20. <https://doi.org/10.1007/s10272-018-0713-z>
- 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>
- Hedström, P. (2006). Explaining Social Change: An Analytical Approach. *Papers: Revista de Sociologia*, 80, 073–95.

- Hedström, P., & Ylikoski, P. (2014). Analytical sociology and rational choice theory. *Analytical Sociology: Norms, Actions and Networks*, 57–70.
- Heilmann, C. (2014). Success conditions for nudges: a methodological critique of libertarian paternalism. *European Journal for Philosophy of Science*, 4(1), 75-94
- Hertwig, R. (2017). When to consider boosting: Some rules for policy-makers. *Behavioural Public Policy*, 1(02), 143–161. <https://doi.org/10.1017/bpp.2016.14>
- Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), 973–986.
- Hertwig, R., & Ryall, M. (2016). Nudge vs. Boost: Agency dynamics under 'libertarian paternalism'. *Boost: Agency Dynamics Under 'Libertarian Paternalism' (January 1, 2016)*.
- Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. a, Kelly, M. P., Nakamura, R., Suhrcke, M., & Ogilvie, D. (2013). Altering micro-environments to change population health behaviour: Towards an evidence base for choice architecture interventions. *BMC Public Health*, 13, 1218. <https://doi.org/10.1186/1471-2458-13-1218>
- Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Lewis, H. B., Wei, Y., Higgins, J. P., & Ogilvie, D. (2015). Portion, package or tableware size for changing selection and consumption of food, alcohol and tobacco. *Cochrane Database of Systematic Reviews*, 9.
- Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80(February), 47–58. <https://doi.org/10.1016/j.socec.2019.03.005>
- Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, 23(1), 1–25.
- Jung, J. Y., & Mellers, B. A. (2016). American attitudes toward nudges. *Judgment and Decision Making*, 11(1).
- Junghans, A. F., Cheung, T. T., & De Ridder, D. D. (2015). Under consumers' scrutiny-an investigation into consumers' attitudes and concerns about nudging in the realm of health behavior. *BMC public health*, 15(1), 336.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Kahneman, D., Slovic, S. P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291. <https://doi.org/10.1111/j.1536-7150.2011.00774.x>
- Katsikopoulos, K. V. (2014). Bounded rationality: The two cultures. *Journal of Economic Methodology*, 21(4), 361–374.
- Keller, C. (2014). Nudging product choices: The effect of position change on snack bar choice. *Food Quality and Preference*, 41(November), 41–43. <https://doi.org/10.1016/j.foodqual.2014.11.005>
- Kelly, J. T. (2013). *Libertarian paternalism, utilitarianism, and justice*.

- Kersbergen, I., Oldham, M., Jones, A., Field, M., Angus, C., & Robinson, E. (2018). Reducing the standard serving size of alcoholic beverages prompts reductions in alcohol consumption. *Addiction*, *113*(9), 1598–1608.
- Kroese, F. M., Marchiori, D. R., & De Ridder, D. T. D. D. (2016). Nudging healthy food choices: A field experiment at the train station. *Journal of Public Health (United Kingdom)*, *38*(2), e133–e137. <https://doi.org/10.1093/pubmed/fdv096>
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478.
- Le Grand, J., & New, B. (2015). *Government paternalism: Nanny state or helpful friend?* Princeton University Press.
- Lee, M., Sul, S., & Kim, H. (2018). Social observation increases deontological judgments in moral dilemmas. *Evolution and Human Behavior*, *39*(6), 611–621.
- Lepenies, R., & Malecka, M. (2015). The Institutional Consequences of Nudging – Nudges, Politics, and the Law. *Review of Philosophy and Psychology*, *6*(427–437).
- Lepenies, R., & Malecka, M. (2019). The ethics of behavioural public policy. *The Routledge Handbook of Ethics and Public Policy*, 513–525. <https://doi.org/10.4324/9781315461731-41>
- Lewis, M. (2016). *The undoing project: A friendship that changed the world*. Penguin UK.
- Lipek, T., Igel, U., Gausche, R., Kiess, W., & Grande, G. (2015). Obesogenic environments: Environmental approaches to obesity prevention. *Journal of Pediatric Endocrinology and Metabolism*, *28*(5–6), 485–495. <https://doi.org/10.1515/jpem-2015-0127>
- Loewe, G. (2006). The development of a theory of rational intertemporal choice. *Papers: Revista de Sociologia*, 195–221.
- Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2015). Warning: You Are About to Be Nudged. In *SSRN*. <https://doi.org/10.2139/ssrn.2417383>
- Loewenstein, G., & Chater, N. (2017). Putting nudges in perspective. *Behavioural Public Policy*, *1*(1), 26.
- Loewenstein, G., & Elster, J. (1992). *Choice over time*. Russell Sage Foundation.
- Loibl, C., Sunstein, C. R., Rauber, J., & Reisch, L. A. (2018). Which Europeans Like Nudges? Approval and Controversy in Four European Countries. *Journal of Consumer Affairs*, *52*(3), 655–688. <https://doi.org/10.1111/joca.12181>
- Lourenço, J. S., Ciriolo, E., Almeida, S. R., & Troussard, X. (2016). Behavioural insights applied to policy: European Report 2016. *Brussels: European Union*.
- Lütge, C., Rusch, H., Uhl, M., & Luetge, C. (2014). *Experimental ethics: Toward an empirical moral philosophy*. Springer.
- Malecka, M. (2020). The normative decision theory in economics: A philosophy of science perspective. The case of the expected utility theory. *Journal of Economic Methodology*, *27*(1), 36–50.
- Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*(6149).
- Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in cognitive sciences*, *22*(4), 280-293

- Meyer, A. (2015). Does education increase pro-environmental behavior? Evidence from Europe. *Ecological Economics*, *116*, 108–121.
- Mills, C. (2013). Why Nudges Matter: A Reply to Goodwin. *Politics*, *33*(1), 28–36. <https://doi.org/10.1111/j.1467-9256.2012.01450.x>
- Mills, C. (2015). The Heteronomy of Choice Architecture. *Review of Philosophy and Psychology*, *6*(March), 495–509. <https://doi.org/10.1007/s13164-015-0242-7>
- Mitchell, G. (2004). Libertarian Paternalism Is an Oxymoron. *Northwestern University Law Review*, *99*, 1245–1276.
- Mongin, P., & Cozic, M. (2018). Rethinking nudge: Not one but three concepts. *Behavioural Public Policy*, *2*(01), 107–124. <https://doi.org/10.1017/bpp.2016.16>
- Mullainathan, S., & Thaler, R. H. (2000). *Behavioral economics* (No. 0898–2937). National Bureau of Economic Research.
- Nagatsu, M. (2015a). Behavioral economics, history of. *International Encyclopedia of the Social & Behavioral Sciences*, *2*, 443–449.
- Nagatsu, M. (2015b). Social Nudges: Their Mechanisms and Justification. *Review of Philosophy and Psychology*, *6*(3), 481–494. <https://doi.org/10.1007/s13164-015-0245-4>
- Newell, R. G. R. G., Siikamäki, J., Siikamaki, J., Siikamäki, J., & Siikamaki, J. (2013). Nudging Energy Efficiency Behavior. *RFF Discussion Paper*, *July*, *44*. <https://doi.org/10.3386/w19224>
- Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and brain sciences*, *37*(1), 1-19.
- Niker, F., Reiner, P. B., & Felsen, G. (2018). Perceptions of Undue Influence Shed Light on the Folk Conception of Autonomy. *Frontiers in Psychology*, *9*, 1400. <https://doi.org/10.3389/fpsyg.2018.01400>
- Ong, H. H., Mullette-Gillman, O., Kwok, K., & Lim, J. (2014). Moral judgment modulation by disgust is bi-directionally moderated by individual sensitivity. *Frontiers in Psychology*, *5*, 194.
- Organisation for Economic Co-operation and Development (OECD). (2017). *Behavioural Insights and Public Policy: Lessons from Around the World*.
- Parzuchowski, M., & Wojciszke, B. (2014). Hand over heart primes moral judgments and behavior. *Journal of Nonverbal Behavior*, *38*(1), 145–165.
- Pérez, C. (2013). Motivational Strategies, Redistributive Policies and Individual Choice. *Working Paper All Theses and Dissertations (ETDs)*, *1332.*, 1–27.
- Petrescu, D. C., Hollands, G. J., Couturier, D., & Ng, Y. (2016). Public Acceptability in the UK and USA of Nudging to Reduce Obesity: The Example of Reducing Sugar-Sweetened Beverages Consumption. *PLoS One*, *11*(6), 1–18. <https://doi.org/10.1371/journal.pone.0155995>
- Pettit, P. (1996). Institutional Design and Rational Choice. In R. E. Goodin (Ed.), *The Theory of Institutional Design* (pp. 54–89). Cambridge University Press.
- Quigley, M. (2013). Nudging for Health: On Public Policy and Designing Choice Architecture. *Medical Law Review*, *21*(4), 588–621. <https://doi.org/10.1093/medlaw/fwt022>
- R Team Core. (2013). *R: A language and environment for statistical computing*.

- Rebonato, R. (2014). A critical assessment of libertarian paternalism. *Journal of Consumer Policy*, 37(3), 357–396. <http://dx.doi.org/10.1007/s10603-014-9265-1>
- Reisch, L. A., & Sunstein, C. R. (2016). Do Europeans like nudges?. *Judgment and Decision making*, 11(4), 310-325.
- Reynolds, J. P., Pilling, M., & Marteau, T. (2018). Communicating quantitative evidence of policy effectiveness and support for the policy: Three experimental studies. *Social Science & Medicine*, 218, 1–12.
- Reynolds, J. P., Stautz, K., Pilling, M., van der Linden, S., & Marteau, T. M. (2020). Communicating the effectiveness and ineffectiveness of government policies and their impact on public support: A systematic review with meta-analysis. *Royal Society Open Science*, 7(1), 190522. <https://doi.org/10.1098/rsos.190522>
- Saghai, Y. (2013). Salvaging the concept of nudge. *Journal of Medical Ethics*, 39(8), 487–493. <https://doi.org/10.1136/medethics-2012-100727>
- Samuels, R., Stich, S., & Bishop, M. (2012). Ending the rationality wars. *Collected Papers, Volume 2: Knowledge, Rationality, and Morality, 1978-2010*, 2, 191.
- Samuelson, P. (1937). A Note on Measurement of Utility. *The Review of Economic Studies*, 4(2), 155–161.
- Schmidt, A. T. (2019). Getting real on rationality—Behavioral science, nudging, and public policy. *Ethics*, 129(4), 511–543.
- Schmidt, A. T., & Engelen, B. (2020). The ethics of nudging: An overview. *Philosophy Compass*, January, 1–13. <https://doi.org/10.1111/phc3.12658>
- Schubert, C. (2015). On the ethics of public nudging: Autonomy and Agency. Available at SSRN 2672970. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2672970](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2672970)
- Sent, E.-M. (2004). Behavioral economics: How psychology made its (limited) way back into economics. *History of Political Economy*, 36(4), 735–760.
- Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, 109(38), 15197–15200.
- Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American Economic Review*, 49(3), 253–283.
- Simon, H. A. (1976). From substantive to procedural rationality. In *25 years of economic theory* (pp. 65–86). Springer.
- Simon, H. A. (1985). Human nature in politics: The dialogue of psychology with political science. *American Political Science Review*, 79(2), 293–304.
- Simon, H. A. (1986). Rationality in psychology and economics. *Journal of Business*, S209–S224.
- Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason (Vol. 3)*. MIT press. MIT press.
- Sims, A., & Müller, T. M. (2019). Nudge versus boost: A distinction without a normative difference. *Economics & Philosophy*, 35(2), 195–222.

- Smith, N. C., Goldstein, D. G., & Johnson, E. J. (2013). Choice Without Awareness: Ethical and Policy Implications of Defaults. *Journal of Public Policy & Marketing*, 32(2), 159–172. <https://doi.org/10.1509/jppm.10.114>
- Sugden, R. (2004). The opportunity criterion: Consumer sovereignty without the assumption of coherent preferences. *American Economic Review*, 94(4), 1014–1033.
- Sugden, R. (2017). Do people really want to be nudged towards healthy lifestyles? *International Review of Economics*, 64(2), 113–123. <https://doi.org/10.1007/s12232-016-0264-1>
- Sunstein, C. R. (2005). Moral heuristics. *Behavioral and brain sciences*, 28(4), 531–541.
- Sunstein, C. R. (2015a). Nudges, agency, and abstraction: A reply to critics. *Review of Philosophy and Psychology*, 6(3), 511–529.
- Sunstein, C. R. (2015b). Nudges Do Not Undermine Human Agency: A Note. In *Journal of Consumer Policy* (forthcoming). <https://doi.org/10.1007/s10603-015-9289-1>
- Sunstein, C. R. (2015c). The Ethics of Nudging. In *Yale Journal on Regulation* (Vol. 32, Issue 2). <http://dx.doi.org/10.2139/ssrn.2526341>
- Sunstein, C. R. (2016). Autonomy by Default. *American Journal of Bioethics*, 16(11), 1–2. <https://doi.org/10.1080/15265161.2016.1222169>
- Sunstein, C. R. (2017a). Nudges that Fail. *Behavioural Public Policy*, 1(1), 4–25.
- Sunstein, C. R. (2017b). People Like Nudges (Mostly). In C. R. Sunstein, *Human Agency and Behavioral Economics* (pp. 17–39). Palgrave Macmillan, Cham. [https://doi.org/10.1007/978-3-319-55807-3\\_2](https://doi.org/10.1007/978-3-319-55807-3_2)
- Sunstein, C. R. (2017c). People Prefer Educative Nudges (Kind Of). In *Human Agency and Behavioral Economics* (pp. 41–72). Palgrave Advances in Behavioral Economics. Palgrave Macmillan, Cham. <https://doi.org/10.1017/CBO9781107415324.004>
- Sunstein, C. R., Reisch, L. A., & Rauber, J. (2018). A worldwide consensus on nudging? Not quite, but almost. *Regulation & Governance*, 12(1), 3–22. <https://doi.org/10.1111/rego.12161>
- 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.
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119(3), 454–458.
- Tannenbaum, D., Fox, C. R., & Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour*, 1(7), 1–7.
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American economic review*, 93(2), 175–179.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Timmons, S., & Byrne, R. M. (2019). Moral fatigue: The effects of cognitive fatigue on moral reasoning. *Quarterly Journal of Experimental Psychology*, 72(4), 943–954.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>

- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Tversky, A., & Kahneman, D. (1996). On the reality of cognitive illusions. *Psychological Review*, 103(3), 582–591.
- Uhlmann, E. L., Pizarro, D. A., & Bloom, P. (2008). Varieties of social cognition. *Journal for the Theory of Social Behaviour*, 38(3), 293–322. <https://doi.org/10.1111/j.1468-5914.2008.00372.x>
- Wansink, B., & Cashman, M. (2006). Mindless eating. The 200 Daily Food Decisions We Overlook. *Environment and Behavior*, 39(1), 106–123. <https://doi.org/10.1177/0013916506295573>
- Wansink, B., Painter, J. E., & North, J. (2005). Bottomless bowls: Why visual cues of portion size may influence intake. *Obesity Research*, 13(1), 93–100. <https://doi.org/10.1038/oby.2005.12>
- Welsh, D. T., & Ordóñez, L. D. (2014). Conscience without cognition: The effects of subconscious priming on ethical behavior. *Academy of Management Journal*, 57(3), 723–742.
- White, M. (2013). *The manipulation of choice: Ethics and libertarian paternalism*. Springer.
- World Bank. (2015). *World development report 2015: Mind, society, and behavior*.
- Yan, H., & Yates, J. F. (2019). Improving acceptability of nudges: Learning from attitudes towards opt-in and opt-out policies. *Judgment and Decision Making*, 14(1), 26.

# Appendix A.

## Supplementary materials for the survey

### 1. Survey and questionnaire

**Table 15. Survey and questionnaire for survey**

Screen and function	Text
1 initial instructions	<p>Next, you will read about different public policy measures.</p> <p>For each measure, you will read a short text that explains its objective and how it works. Then you will answer some questions:</p>
Start of the question about nudges. Subjects saw questions in random order. Screen 2 to 18.	
Default organ donation	<p>In many countries, there is a shortage of organ donors to meet the demands of patients who need a transplant. To increase the number of donors, the administration promotes a measure according to which the entire adult population is considered a default donor. Those who do not wish to be a donor can drop out by actively expressing it. The adoption of default options has been found to have a significant influence on decisions. In particular, people tend to stick with the default option, and thus the measure increases the number of donors.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Organ donation campaign	<p>In many countries, there is a shortage of organ donors to meet the demands of patients who need a transplant. To increase the number of donors, the administration is promoting a campaign designed to sensitise citizens about the importance of becoming a donor. The campaign highlights that Spain is the European leader in organ donation and calls on citizens to get a donor card.</p> <p>It has been found that reporting the behaviour of the majority increases people's motivation to engage in such behaviour; in this case, it increases the probability that they will become donors.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p>

	<p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Default energy use	<p>To reduce excessive energy consumption in hotels, the administration is promoting a regulation according to which, by default, hotels will not renew guest towels daily. If guests want clean towels daily, they can ask for it, but they will not receive them if there is no active demand on their part. Adopting this default option makes it less likely that people will demand daily towel cleaning and contributes to energy savings.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Social comparison of energy use	<p>The excessive consumption of energy in homes leads to a waste of economic resources for the home and generates environmental problems. To reduce energy consumption, the administration promotes a new electricity bill. The new bill includes information that compares the energy consumption of the home that receives the bill with the consumption of homes in their neighbourhood of similar size. If the household consumption is lower, a positive message is indicated. If the energy consumption of the home is higher, the letter indicates it and provide simple guidelines to reduce energy consumption. Using social comparison has been found to influence people's behaviour and increase their motivation to save energy.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Default donations to charity	<p>To increase charitable donations, the administration has decided to change the income tax form and preselect by default the option to donate. People who do not wish to donate can modify the selection. Adopting this default option makes it more likely that people will donate to charities and the third sector.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p>

	<p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Charitable donations campaign	<p>To increase donations to charity and the third sector, coinciding with the start of the tax collection campaign, the administration launches an awareness campaign that explains the work of these entities to convey to citizens the importance of donating to these associations. The use of emotional messages has been found to influence decisions and can be helpful in motivating people to donate to these entities.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Carbon reduction app	<p>To reduce pollution derived from the use of fossil fuel transport, the administration plans to create an app linked to Google Maps that to the information on route options, duration and type of transport adds, for each type of transport, its environmental impact. The objective of the measure is to inform and motivate users to choose the least polluting transport.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Optical illusions to control speeding	<p>To reduce car accidents caused by speeding, the administration will paint white stripes on the asphalt of some roads that are progressively narrow. These stripes have been found to create an optical illusion of increased speed and induce the driver to slow down their vehicle.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p>

	<p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Campaign to control speeding	<p>To reduce car accidents caused by speeding, the administration is promoting an awareness campaign with messages that refer to the damages and victims caused by drivers who exceed the speed limit. The use of emotional messages has been found to influence decisions and can help motivate drivers to be more cautious behind the wheel.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Tax compliance campaign	<p>To avoid tax fraud, the administration is promoting a campaign that uses messages such as "paying taxes is the right decision" and "most taxpayers pay their taxes and contribute to the common good." The measure intends to generate increase the motivation to comply.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Default food layouts	<p>To promote a healthy diet among the population, the administration plans to implement a measure to modify the disposition of products in supermarkets and restaurants. Healthy products will be placed in more visible and easily accessible places and unhealthy products will be located in places that are less visible and difficult to access. The position of products has been found to affect people's decisions. Those nearby and accessible are chosen more; therefore, the measure is useful to encourage the consumption of healthier options.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p>

	<p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Traffic light labels	<p>To promote a healthy diet among citizens, the administration plans to implement traffic light labels on the packaging of different products. Unhealthy products (high in fat and sugar and low in vitamins and minerals) will be marked with a red label; healthy products (high in vitamins and minerals and low in fat and sugar) will be marked with a green label. By association with the traffic light, the labels inform about the product and are designed to motivate healthier choices.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Automatic enrolment for social benefits	<p>There is evidence that indicates that there are people who are entitled to receive a social benefit but do not apply for it. To avoid this situation, the administration is promoting a measure to automatically pay people the benefits and deductions that they are eligible for.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Gambling control app	<p>To reduce the problems of addiction due to online gambling, the administration promotes a measure according to which users of gambling apps and websites must establish a limit of use and a maximum limit of bets per week. Once the limit is exceeded, the applications will block user access. Users will keep the option to manage their limits and modify them if they wish, as long as they have not exceeded their limit in that week. Establishing self-regulation strategies is effective in avoiding impulsive decisions and controlling addictive behaviour. With this measure, the administration wants to reduce and avoid cases of addiction to gambling.</p>

	<p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Gambling control campaign	<p>To reduce the problems of addiction due to online gambling, the administration is promoting an awareness campaign to make the population aware of the problems associated with gambling addiction, highlighting aspects such as the high probability of losing money and psychological problems associated with the addiction. It has been found that pointing out the risks and losses associated with certain behaviours influences individual decisions and can be useful in reducing and avoiding cases of gambling addiction.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Regulation of alcohol container sizes	<p>To reduce alcohol consumption among the population, the administration is promoting a measure to regulate the size of the servings of alcoholic beverages (beer and wine) that are served in bars and restaurants and to make the containers and glasses smaller. Reducing the size makes people more likely to consume less and helps reduce alcohol consumption among the population.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
Voting reminders	<p>To increase electoral participation, a few days before the elections, the administration encourages the sending of reminders to citizens via SMS. The message includes basic information about the elections, for example, the time, place and necessary documentation, and encourage to saved in the calendar an event: "Go to vote". Reminders are informative, help people make concrete plans</p>

	<p>to vote, and make them more likely to stick with them. With this measure, the administration can expect an increase in electoral participation.</p> <p>Do you agree with the application of this measure? (1 = No, nothing; 5 = Yes, totally)</p> <p>1_2_3_4_5</p> <p>Do you think this measure will be effective? (1 = No, not at all effective, 5 = Yes, totally effective)</p> <p>1_2_3_4_5</p> <p>Do you consider that this measure manipulates behaviour? (1 = No, not at all manipulative; 5 = Yes, totally manipulative)</p> <p>1_2_3_4_5</p>
End of the survey. Subjects start the questionnaire. Screens 18 to 37	
Nationality	<p>What is your Nationality?</p> <ul style="list-style-type: none"> <li><input type="radio"/> Spanish</li> <li><input type="radio"/> Spanish and other_____</li> <li><input type="radio"/> Foreign</li> </ul>
Activity	<p>What is your main activity?</p> <ul style="list-style-type: none"> <li><input type="radio"/> I work full time</li> <li><input type="radio"/> Part-time work</li> <li><input type="radio"/> Retiree or pensioner</li> <li><input type="radio"/> Unemployed (I have worked before)</li> <li><input type="radio"/> Unemployed (I'm looking for my first job)</li> <li><input type="radio"/> Student</li> <li><input type="radio"/> Housework</li> </ul>
Studies	<p>What is your level of completed studies?</p> <ul style="list-style-type: none"> <li><input type="radio"/> Without studies (Primary studies without finishing)</li> <li><input type="radio"/> First Grade (School certificate, EGB 1st stage, more or less ten years)</li> <li><input type="radio"/> Second Grade. 1st Cycle (School graduate, or EGB 2nd stage, 1st and 2nd ESO-1st cycle- up to 14 years)</li> <li><input type="radio"/> Second Grade. 2nd Cycle (FP I and II, Higher Bachelor, BUP, 3rd and 4th ESO (2nd cycle) COU, PREU, 1st and 2nd Bachillerato, up to 18 years)</li> <li><input type="radio"/> Third Grade. 1st Cycle (Equivalent to Technical Engineer, 3 years, University Schools, Technical Engineers, Technical Architects, Experts, Teaching, ATS, University Graduates, 3-year career, Social graduates, Social workers, etc.)</li> <li><input type="radio"/> Degree, Degree. 2nd Cycle (University, Higher Graduates, Faculties, Higher Technical Schools, etc.</li> <li><input type="radio"/> Third Degree (Master)</li> <li><input type="radio"/> Third grade (Doctorate)</li> </ul>
Religion	<p>How do you define yourself in religious matters?</p> <ul style="list-style-type: none"> <li><input type="radio"/> Catholic / Christian</li> <li><input type="radio"/> Believer in another religion, which one? _____</li> <li><input type="radio"/> Agnostic or atheist</li> </ul>
Ideology	<p>When talking about politics, the expressions "left" and "right" are commonly used. Where would you place your opinion / political position? (0 = extreme left; 10 = extreme right)</p> <p> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5    <input type="radio"/> 6    <input type="radio"/> 7    <input type="radio"/> 8    <input type="radio"/> 9    <input type="radio"/> 10 </p>

	<p>or</p> <p>I prefer not to answer</p>
Trust	<p>Do you trust the government and the public administration to develop politics to promote social wellbeing? (1 = Not at all, 5 = Yes, totally)</p> <p style="text-align: center;"> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5 </p>
Support for government interference	<p>Are you in favour of the administration trying to change the behaviour of individuals in order to improve social welfare? (1 = Not at all, 5 = Yes, totally)</p> <p style="text-align: center;"> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5 </p> <p>What if it were individual wellbeing? (1 = Not at all, 5 = Yes, totally)</p> <p style="text-align: center;"> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5 </p>
Organ donor	<p>Are you an organ donor?</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p>
Concern environment	<p>Do you care about protecting the environment?</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p>
Charity donation	<p>Do you make or have you made donations to charities or the third sector?</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p>
Type of transportation	<p>To move around in your day to day, what type of transport do you use the most?</p> <p><input type="radio"/> Private car</p> <p><input type="radio"/> Private motorcycle</p> <p><input type="radio"/> Cab</p> <p><input type="radio"/> Public transport (bus, metro, train, etc.)</p> <p><input type="radio"/> Bike</p> <p><input type="radio"/> Walking</p> <p><input type="radio"/> Other</p>
Traffic fines	<p>Have you ever had a ticket or for speeding?</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> No, I don't drive, and I don't have a driver's license</p>
Social benefits	<p>Are you a beneficiary of a social benefit?</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p>
Tax	<p>Do you file your income tax annually?</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No, I do not have to do it</p> <p><input type="radio"/> No, I've only done it a few years</p> <p><input type="radio"/> No, it is made for me by a relative</p> <p><input type="radio"/> No, it is made for me by a professional</p>
Gambling	<p>Do you gamble on apps or on online?</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No, but have I ever played</p>

	<input type="radio"/> No, I have never played
Vote	Do you vote in the elections? <input type="radio"/> Yes, always <input type="radio"/> Yes, I have voted, but I don't always vote <input type="radio"/> Never <input type="radio"/> No, I can't vote
Type of diet	Do you consider that you follow a healthy diet? <input type="radio"/> Yes, mostly <input type="radio"/> Neither healthy nor unhealthy <input type="radio"/> No
Help to eat healthily	Do you think you could benefit from some external help to eat healthier? <input type="radio"/> Yes <input type="radio"/> No
Alcohol consumption	How often do you drink alcoholic (wine and/or beer)? <input type="radio"/> Never <input type="radio"/> Rarely <input type="radio"/> Once a month <input type="radio"/> Once every two weeks <input type="radio"/> Once a week <input type="radio"/> 2 to 4 times a week <input type="radio"/> More than 4 times a week
Situationism self	Do you consider that your behaviour is mainly determined... (1 = By my choices and decisions, or 5 = By the context and circumstances that surround me) <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5
Situationism others	Do you consider that the behaviour of the majority of people is mainly determined... (1 = By their choices and decisions, 5 = By the context and circumstances that surround them) <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5

## 2. Data preparation for the survey

Specific variables had to be transformed to meet the requirements of robustness before proceeding with the statistical analysis. To analyse the relationship between categorical variables with the chi-square test several of the variable's categories were combined, increasing the number of observations in the subgroups and ensuring that none of the categories' expected frequencies were lower than 5%. The recoded variables were also used for the logistic regression.

*Education attainment* was measured using eight categories, yet the number of observations in some of the categories was too small. To ensure statistical relevance, the eight original categories were combined into four categories: (1) less than secondary studies, combining the categories: unfinished education or non and primary first and second stage school certificate; (2) secondary studies, including the original category second cycle secondary degree education, (3) university studies, combining two categories of university education of first and second cycle, and finally (4) master's degree and PhD education graduates.

*Activity* was initially measured using seven categories; however, the category 'unemployed looking for first job' was residual. To ensure statistical relevance, it was combined with the category unemployed, creating a single category for those that do not work but are looking for a job. Activity had an abnormal distribution, 50% of the sample is working full-time, and the rest of the categories are less frequent. It is, however, not possible to group categories because it does not make theoretical sense. To test the effect of being employed, I created a new variable with two categories (1) working, including those who work full time and part-time and (2) not working, including those who are unemployed, studying or looking for a job. In the cases in which it was of interest to understand how specific activity groups, such as students and those who do housework react to nudges, the non-parametric Fisher's exact test has been used to test the relation between variables. The variable is excluded from the regression models because it was not significantly related with the 17 dependent variables and the non-working categories had very few cases.

*Region (CCAA)* was not included in the analysis because the sample was too small to find differences among regions.

*Religion* was initially measured using three categories and allowed a non-response: Catholic/Christian, Other religion, Agnostic and Atheists. The percentage of people in the category 'other religion' was very low, to be included in the analysis those that declare themselves as religious were all collapsed in a single category: religious. Respondents that choose not to answer were treated as missing values.

The variable *Nationality* was discharged for the analyse because only 2% of the sample did not have the Spanish Nationality, with such level it failed to make an impact on the models.

*Ideology* was measured initially with a 0 to 10 Likert-type scale. Due to the small number of observations in some categories, I gathered some levels: 0,1,2,3 responses were categorised as 'left', 4, 5, 6 responses were categorised as 'centre', 7,8,9,10 responses were categorised as 'right'. Respondents that choose not to answer were treated as missing values.

*Trust in institutions*, *Support for government interference with individual choices for social good*, and *Support for government interference with individual choices for own good* were measured using a 1 to 5 Likert-scale, from no trust/strongly disagree to strong trust/strongly agree. To ensure the robustness of the statistical analysis, the variables were recorded into three categories: 'no-trust in institutions' and 'no support for government interference for individual good' and 'no support for government interference for social good', combining 1 and 2 response, neutral including three response and 'trust in institutions', 'support for government interference for individual good' and 'support for government interference for social good' combining 4 and 5 responses.

Likewise, self-perception of dispositionism vs situationism and perception of others dispositionism vs situationism were measured using a 1 to 5 Likert-scale: my/others behaviour is mainly determined 1 = my/others choices and decisions, five = the context and the circumstances that surround me/others. To ensure the robustness of the statistical analysis, the variables were recorded into three categories: 1 and 2 responses were categorised as 'dispositionism', three was considered neutral, 4 and 5 responses were categorised as 'situationism'.

Finally, a few of the behavioural factors that were included in the regressions were also categorised to avoid categories with few cases:

- *Type of regular form of transportation* was initially measured using seven categories: car, motorcycle, taxi, walking, public transportation (including bus, metro and taxi), bicycle and others. Due to lack of cases in some categories, the variable was dichotomised into two categories: (1) polluting transport, including the categories car, motorcycle and taxi and (2) non-polluting transport, including, the use of public transportation, walking and bicycles.
- *Traffic fines* measured whether people had been fined before using three categories: fined before, not fined before, and no drivers. For the analysis, the variable was dichotomised into two categories: (1) fined, and (2) non-fined, combining those that declared not having been fined before and non-drivers.
- *Taxpayer* measured whether respondents pay or had paid taxes before, and who did the tax form. The original measures were collapsed into three categories: (1) taxpayers who completed the tax forms by themselves, (2) taxpayers who have someone else complete their tax forms and (3) non-taxpayers.
- *Gamblers* measured whether respondents regularly gamble. Measured initially in three categories (regular gambler, occasional gambler, and non-gambler), the variable was

dichotomised into two categories, (1) gamblers, including regular and occasional gamblers and (2) non-gamblers.

- *Voters* measured whether respondents regularly voted in elections. The variable had four categories (always votes, occasionally votes, has never voted before, cannot vote). Due to lack of cases in the non-voting categories, the variable was dichotomised into two categories: (1) always votes and (2) not-always votes, including the categories occasional voters, no-voters and not voters yet.
- *Alcohol consumption* collected respondent's self-reported alcohol consumption using six categories: never drinks, rarely drinks, drinks once a month, drinks once every two weeks, drinks once every week, drinks from two to fourth times a week, drinks more than forth times a week. These seven categories were transformed into four: (1) non-drinkers, including the never, drinks and rarely drinks categories; (2) moderate consumption, including the categories drinks once a month, drinks once every two weeks, (3) regular consumption, including the categories drinks once every week, drinks from two to fourth times a week, and (4) high consumption including the category drinks more than forth times a week.

### 3. Co-variables descriptive statics

**Table 16. Sample characteristics of participants in the survey**

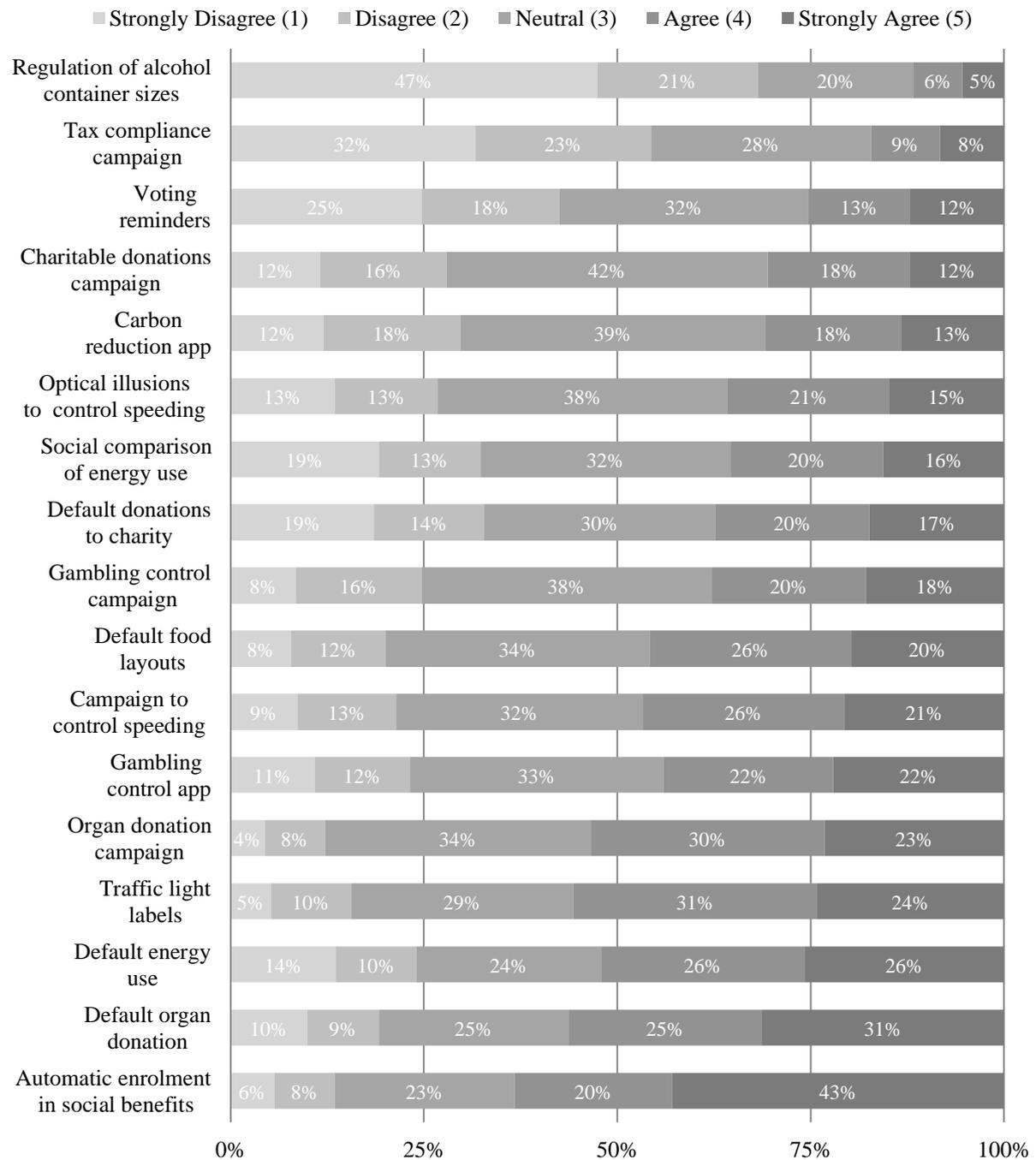
Factors	Distribution of respondents
Age	Range:18-75 Median: 44 Mean + sd: 45.11, 15.3
Biological sex	Male: 48.78% Female: 51.21%
Highest education level	Unfinished education or non: 0.96% School: 2.10% High school (1): 7.91% High school (2): 38.28% College (1): 19.87% College or university degree (2): 20.19% Port-graduate degree (master's): 8.72% Post-graduate degree (PhD): 1.93%
Activity	Full-time job: 50.08% Part-time job: 10.33% Retired: 16.63% Unemployed: 8.23% Unemployed (fist job): 0.64% Student: 10.01% Housework: 4.03%
Region (CCAA)	Andalucía: 18.57% Aragón: 2.90% Principado de Asturias: 2.26% Illes Balears: 1.29% Canarias: 4.84 % Cantabria: 1.93% Castilla y león: 7.59% Castilla-la Mancha: 4.36% Catalunya: 16.63% Comunitat Valenciana:11.63% Extremadura: 0.80% Galicia: 5.16% Madrid: 13.89% Murcia: 2.26% Navarra: 1.13% País Vasco: 3.87% La Rioja: 0.80% Ceuta: 0% Melilla: 0%
Nationality	Spanish: 96.44% Spanish and other: 1.29%

	Others: 2.26%
Religion	Catholic/Christian: 42.81% Other religions: 2.10% Agnostic and atheist: 42.81% Na: 12.27%
Ideology	(0=extreme left, 10= extreme right) 0: 1.93% 1: 4.20% 2: 13.08% 3: 16.15% 4: 10.33% 5: 17.77% 6: 7.91% 7: 7.43% 8: 4.20% 9: 1.13% 10: 0.64% Na: 15.18%
Trust	Not at all: 20.03% No: 21.16% Neutral: 40.71% Yes: 12.76% Very much so: 5.33%
Government interference for social good	Not at all: 8.07% No: 10.82% Neutral: 31.50% Yes: 29.56% Very much so: 20.03%
Government interference for individual good	Not at all: 9.04% No: 11.63% Neutral: 31.74% Yes: 24.39% Very much so: 15.18%
Organ donors	Yes: 32.31 % No: 67.68%
Environmental consciousness	Environmental conservation as a priority: 78.51% No concern environmental conservation: 21.48%
Charity donations	Doing or having made charity donations: 61.06% Not having made charity donations: 38.93%
Regular form of transportation	Car: 41.84% Motorcycle: 4.36% Taxi: 0.16% Public transportation (bus, metro, train): 21.64% Bicycle: 2.10% Walking: 29.24% Others: 0.64%

Traffic fines	Has had traffic fines: 27.46% No traffic fines: 64.78% Does not drive: 7.75%
Social befits	Receives social benefit: 10.32% Does not receives social benefits: 89.66%
Taxpayer	Taxpayer that makes their tax declaration: 73.50% No taxpayer: 15.50% Occasional taxpayer: 2.26% Relative does the taxpayer's declaration: 4.36% A financial manager makes the taxpayer's declaration: 4.36%
Gamblers	Regular gambler: 1.45% Occasional gambler: 12.43% Does not gamble: 86.10%
Voters	Always vote during elections: 80.12% Occasional voter: 16.47% Never voted before: 2.58% Cannot vote: 0.80%
Self-reported diet	Declares maintaining a healthy diet: 55.25% Declares maintaining a somewhat healthy diet: 38.77% Declares having an unhealthy diet: 5.97%
Diet beliefs	Declares could benefit from external help to improve their diet: 64.78% Declares not wanting external help to improve their diet: 35.21%
Alcohol consumption	Never: 15.99% Rarely: 17.12% Once a month: 9.69% Once every two weeks: 7.91% Once a week: 20.03% From twice to forth times a week:16.47% More than forth times a week:12.76%
Self-perception of dispositionism vs situationism	My behaviour is mainly determined (1 = by my choices and decisions, 5 = by the context and the circumstances that surround me) 1: 24.71% 2: 24.39% 3: 36.51% 4: 7.91% 5: 6.46%
Others perception of dispositionism vs situationism	The behaviour of most people is mainly determined (1 = by their choices and decisions, 5 = by the context and circumstances around them) 1: 6,78% 2: 7.59% 3: 37.96% 4: 27.30% 5: 20.35%

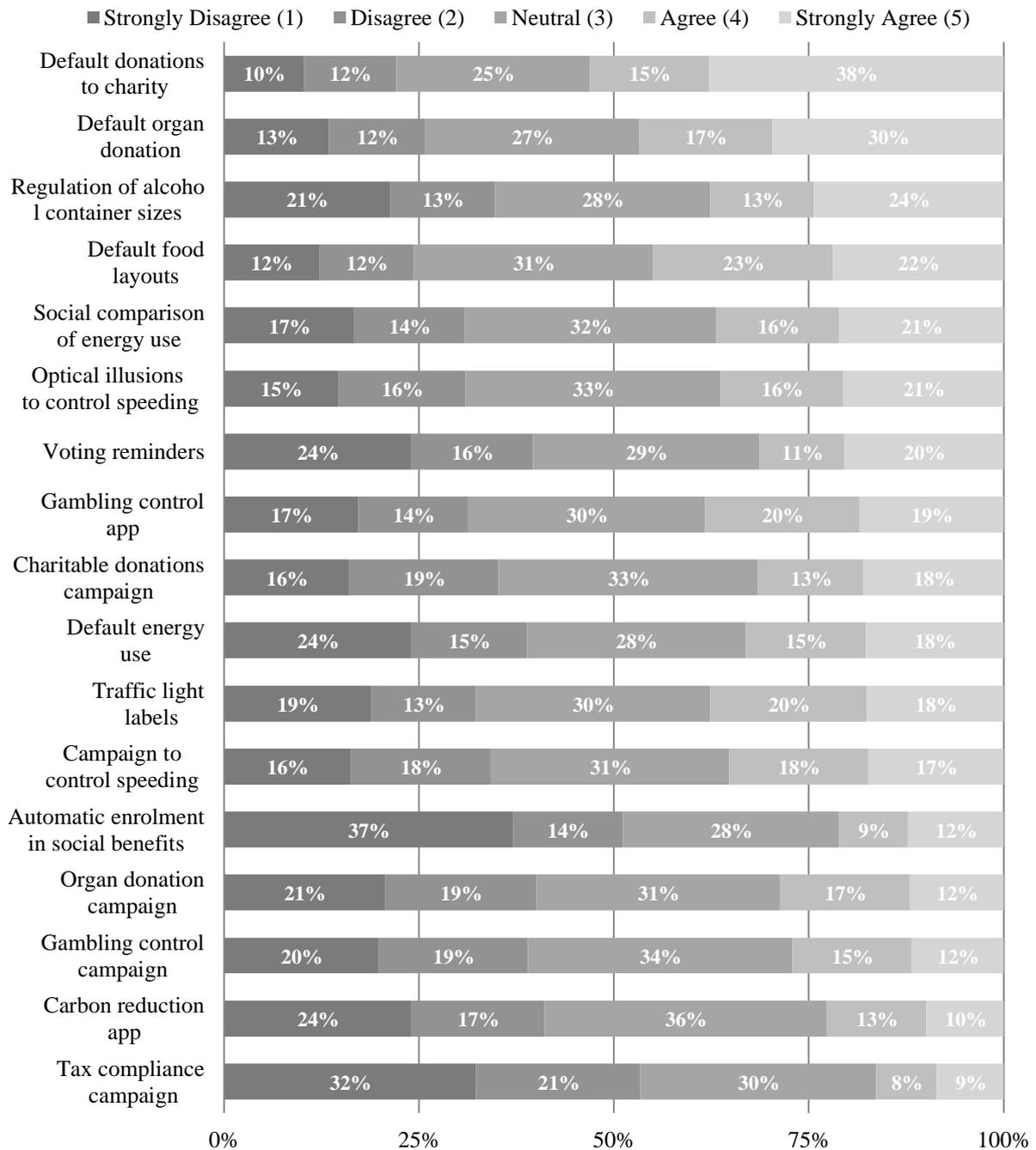
#### 4. Data on perceived effectiveness and perceived manipulation

Figure 7. Perceived effectiveness for the 17 nudges.



Percentages for each category of the five point-Likert scale. Nudges ranked by 'Strongly Agree' responses, from the nudge perceived as less effective to the one perceived as more effective.

**Figure 8. Perceived manipulation for the 17 nudges.**



Percentages for each category of the five point-Likert scale. Nudges ranked by 'Strongly Agree' responses, from the nudge perceived as more manipulative to the one perceived as less manipulative.

# Appendix B.

## 1. Supplementary materials Study A.

### 1.1. Survey and questionnaire of Study A.

**Table 17. Experimental sequence and questions Study A.**

Screen and function	Text
1 Initial instructions in experimental C0, C2, C5	<p>In Spain, the obesity rate has doubled in the last 20 years. According to the National Institute of Statistics, 36% of the adult population is overweight, and 17% is obese. Both conditions harm individual health and multiply the risk of mortality from cardiovascular problems, diabetes and cancer. Also, they create social problems and have very significant economic health cost.</p> <p>The consumption of sugary drinks contributes significantly to this problem. These products contain high amounts of sugar and calories and have little nutritional value.</p> <p>To address this problem, the administration is considering <u>applying different measures to reduce the consumption of sugary drinks among the population.</u></p> <p>Next, you will see eight measures to meet this goal. Then you will answer some questions.</p>
1 Initial instructions in experimental conditions C1, C2, C4	<p>In Spain, the obesity rate has doubled in the last 20 years. According to the National Institute of Statistics, 36% of the adult population is overweight, and 17% is obese. Both conditions harm individual health and multiply the risk of mortality from cardiovascular problems, diabetes and cancer. Also, they generate social problems and have very significant economic health cost.</p> <p>The consumption of sugary drinks contributes significantly to this problem. These products contain high amounts of sugar and calories and have little nutritional value.</p> <p>To address this problem, the administration is considering <u>applying different measures to reduce the consumption of sugary drinks among the population.</u></p> <p>Next, you will see eight measures to meet this goal. Then you will answer some questions.</p>
2 Initial instructions in experimental conditions C1, C2, C4	<p>Attention!</p> <p>To review each measure, you have a limited time of 40 seconds. Do not worry, you have time to read and answer, but you must be focused. At the top left of your screen have a counter that tells you how much time you have.</p>
Start the main survey. Screens 2/3 to X). Rotation between questions.	
<p><b>Size</b></p> <p>C0 C1</p>	<p>Reducing the sugar-sweetened beverages containers to smaller sizes (price adjustment to new size).</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;">○ ○ ○ ○ ○ ○</p>

		(Not at all) 0 1 2 3 4 5 (Yes, totally)
	C2 C3	<p>Reducing the sugar-sweetened beverages containers to smaller sizes (price adjustment to new size).</p> <p>The measure has been found to be effective, and significantly reduces calorie consumption.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;"> <input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5 </p> <p>(Not at all) (Yes, totally)</p>
	C4 C5	<p>Reducing the sugar-sweetened beverages containers to smaller sizes (price adjustment to new size).</p> <p>It has been found that this measure automatically and subconsciously reduces calorie consumption.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;"> <input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5 </p> <p>(Not at all) (Yes, totally)</p>
<b>Location</b>	C0 C1	<p>Changing the placement of sugary drinks in supermarkets and restaurants so that they are placed in places that are less visible and difficult to access.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;"> <input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5 </p> <p>(Not at all) (Yes, totally)</p>
	C2 C3	<p>Changing the placement of sugary drinks in supermarkets and restaurants so that they are placed in places that are less visible and difficult to access.</p> <p>The measure has been found to be effective, significantly reducing the choice of these products and, therefore, calorie consumption.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;"> <input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5 </p> <p>(Not at all) (Yes, totally)</p>
	C4 C5	<p>Changing the placement of sugary drinks in supermarkets and restaurants so that they are placed in places that are less visible and difficult to access.</p> <p>It has been found that this measure automatically and subconsciously reduces the choice of these products and, therefore, the consumption of calories.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;"> <input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5 </p>

		(Not at all) 0 1 2 3 4 5 (Yes, totally)
<b>Default</b>	C0 C1	<p>Establishing water as the drink included in restaurant menus by default.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;">○ ○ ○ ○ ○ ○</p> <p>(Not at all) 0 1 2 3 4 5 (Yes, totally)</p>
	C2 C3	<p>Establishing water as the drink included in restaurant menus by default.</p> <p>The measure has been found to be effective, significantly reducing the choice of sugary drinks and, therefore, calorie consumption.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;">○ ○ ○ ○ ○ ○</p> <p>(Not at all) 0 1 2 3 4 5 (Yes, totally)</p>
	C4 C5	<p>Establishing water as the drink included in restaurant menus by default.</p> <p>It has been found that this measure automatically and subconsciously reduces the choice of these products and, therefore, the consumption of calories.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;">○ ○ ○ ○ ○ ○</p> <p>(Not at all) 0 1 2 3 4 5 (Yes, totally)</p>
<b>Labels</b>	C0 C1	<p>Introducing traffic light like labels on the packaging of different products and mark high sugar beverages in red to highlight that they are unhealthy.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;">○ ○ ○ ○ ○ ○</p> <p>(Not at all) 0 1 2 3 4 5 (Yes, totally)</p>
	C2 C3	<p>Introducing traffic light like labels on the packaging of different products and mark high sugar beverages in red to highlight that they are unhealthy.</p> <p>The measure has been found to be effective, significantly reducing the choice of these products and, therefore, calorie consumption.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;">○ ○ ○ ○ ○ ○</p> <p>(Not at all) 0 1 2 3 4 5 (Yes, totally)</p>



		(Not at all)	0	1	2	3	4	5	(Yes, totally)
	C2 C3	Developing an App that allows users to control their consumption of sugary drinks, notify them in case of excessive consumption and recommend alternative products.							
		The measure has been found to be effective, significantly reducing the choice of these products and, therefore, calorie consumption.							
		Do you agree with the implementation of this measure?							
			<input type="radio"/>						
		(Not at all)	0	1	2	3	4	5	(Yes, totally)
	C4 C5	Developing an App that allows users to control their consumption of sugary drinks, notify them in case of excessive consumption and recommend alternative products.							
		It has been found that this measure consciously reduces the choice of these products and, therefore, calorie consumption.							
		Do you agree with the implementation of this measure?							
			<input type="radio"/>						
		(Not at all)	0	1	2	3	4	5	(Yes, totally)
<b>Tax</b>	C0 C1	Introducing a tax on sugary drinks to increase their purchase price.							
		Do you agree with the implementation of this measure?							
			<input type="radio"/>						
		(Not at all)	0	1	2	3	4	5	(Yes, totally)
	C2 C3	Introducing a tax on sugary drinks to increase their purchase price.							
		The measure has been found to be effective, significantly reducing the choice of these products and, therefore, calorie consumption.							
		Do you agree with the implementation of this measure?							
			<input type="radio"/>						
		(Not at all)	0	1	2	3	4	5	(Yes, totally)
	C4 C5	Introducing a tax on sugary drinks to increase their purchase price.							
		It has been found that this measure consciously reduces the choice of these products and, therefore, calorie consumption.							
		Do you agree with the implementation of this measure?							
			<input type="radio"/>						
		(Not at all)	0	1	2	3	4	5	(Yes, totally)

<b>Subliminal</b>	C0 C1	<p>Playing subliminal messages in movie and television commercials to induce the consumption of sugar-free and healthier drinks.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;"> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5    (Yes, totally) </p> <p>(Not at all)</p>
	C2 C3	<p>Playing subliminal messages in movie and television commercials to induce the consumption of sugar-free and healthier drinks.</p> <p>The measure has been found to be effective, significantly reducing the choice of sugary drinks and, therefore, calorie consumption.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;"> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5    (Yes, totally) </p> <p>(Not at all)</p>
	C4 C5	<p>Playing subliminal messages in movie and television commercials to induce the consumption of sugar-free and healthier drinks.</p> <p>It has been found that this measure automatically and subconsciously reduces the choice of these products and, therefore, the consumption of calories.</p> <p>Do you agree with the implementation of this measure?</p> <p style="text-align: center;"> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5    (Yes, totally) </p> <p>(Not at all)</p>
Screen 9/10 to 17/18		
<b>Qualitative question</b>	all	<p>What has been the most important thing for you when assessing the acceptance of the different measures?</p> <div style="border: 1px solid black; height: 50px; width: 100%;"></div>
<b>Size</b>	C0 C1	<p>Regarding the measure of limiting the size of the packaging of sugary drinks:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;"> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5    (Yes, totally effective) </p> <p>(Not effective at all)</p> <p>Do you think it manipulates behaviour?</p> <p style="text-align: center;"> <input type="radio"/> 0    <input type="radio"/> 1    <input type="radio"/> 2    <input type="radio"/> 3    <input type="radio"/> 4    <input type="radio"/> 5    (Yes, totally manipulative) </p> <p>(Not manipulative at all)</p>
	C2 C3	<p>Regarding the measure of limiting the size of the packaging of sugary drinks:</p>

		<p>Do you think it manipulates behaviour?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, totally manipulative at   all)   manipulative)</p>
	C4 C5	<p>Regarding the measure of limiting the size of the packaging of sugary drinks:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, effective at   all)   totally effective)</p>
<b>Location</b>	C0 C1	<p>Regarding the measure of placing sugary drinks in places that are less visible and difficult to access:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, effective at   all)   totally effective)</p> <p>Do you think it manipulates behaviour?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, totally manipulative at   all)   manipulative)</p>
	C2 C3	<p>Regarding the measure of placing sugary drinks in places that are less visible and difficult to access:</p> <p>Do you think it manipulates behaviour?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, totally manipulative at   all)   manipulative)</p>
	C4 C5	<p>Regarding the measure of placing sugary drinks in places that are less visible and difficult to access:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, effective at   all)   totally effective)</p>
<b>Default</b>	C0 C1	<p>Regarding the measure of establishing water as the default drink included in restaurant menus:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, effective at   all)   totally effective)</p> <p>Do you think it manipulates behaviour?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, totally manipulative at   all)   manipulative)</p>
	C2 C3	<p>Regarding the measure of establishing water as the default drink included in restaurant menus:</p>

		<p>Do you think it manipulates behaviour?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, totally manipulative at   all)   manipulative)</p>
	C4 C5	<p>Regarding the measure of establishing water as the default drink included in restaurant menus:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, effective at   all)   totally effective)</p>
<b>Labels</b>	C0 C1	<p>Regarding the measure of introducing traffic-like labels on high sugar beverages:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, effective at   all)   totally effective)</p> <p>Do you think it manipulates behaviour?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, totally manipulative at   all)   manipulative)</p>
	C2 C3	<p>Regarding the measure of introducing traffic-like labels on high sugar beverages:</p> <p>Do you think it manipulates behaviour?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, totally manipulative at   all)   manipulative)</p>
	C4 C5	<p>Regarding the measure of introducing traffic-like labels on high sugar beverages:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, effective at   all)   totally effective)</p>
	C0 C1	<p>Regarding the measure to promote an awareness campaign on the problems associated with the consumption of sugary drinks:</p> <p>Do you think it will be effective?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, effective at   all)   totally effective)</p> <p>Do you think it manipulates behaviour?</p> <p style="text-align: center;">○   ○   ○   ○   ○   ○</p> <p>(Not   0   1   2   3   4   5   (Yes, totally manipulative at   all)   manipulative)</p>
	C2 C3	<p>Regarding the measure to promote an awareness campaign on the problems associated with the consumption of sugary drinks:</p>

		<p>Do you think it manipulates behaviour?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5   (Yes, totally manipulative)</p> <p>(Not manipulative at all)</p>
	C4 C5	<p>Regarding the measure to promote an awareness campaign on the problems associated with the consumption of sugary drinks:</p> <p>Do you think it will be effective?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5   (Yes, totally effective)</p> <p>(Not effective at all)</p>
<b>App</b>	C0 C1	<p>Regarding the measure of developing an application to self-regulate the consumption of sugary drinks:</p> <p>Do you think it will be effective?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5   (Yes, totally effective)</p> <p>(Not effective at all)</p> <p>Do you think it manipulates behaviour?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5   (Yes, totally manipulative)</p> <p>(Not manipulative at all)</p>
	C2 C3	<p>Regarding the measure of developing an application to self-regulate the consumption of sugary drinks:</p> <p>Do you think it manipulates behaviour?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5   (Yes, totally manipulative)</p> <p>(Not manipulative at all)</p>
	C4 C5	<p>Regarding the measure of developing an application to self-regulate the consumption of sugary drinks:</p> <p>Do you think it will be effective?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5   (Yes, totally effective)</p> <p>(Not effective at all)</p>
<b>Tax</b>	C0 C1	<p>Regarding the measure of introducing a tax on sugary drinks:</p> <p>Do you think it will be effective?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5   (Yes, totally effective)</p> <p>(Not effective at all)</p> <p>Do you think it manipulates behaviour?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5   (Yes, totally manipulative)</p> <p>(Not manipulative at all)</p>
	C2 C3	<p>Regarding the measure of introducing a tax on sugary drinks:</p> <p>Do you think it manipulates behaviour?</p> <p><input type="radio"/> 0   <input type="radio"/> 1   <input type="radio"/> 2   <input type="radio"/> 3   <input type="radio"/> 4   <input type="radio"/> 5</p>

		(Not manipulative at all)	0	1	2	3	4	5	(Yes, totally manipulative)
	C4 C5	Regarding the measure of introducing a tax on sugary drinks:							
<b>Subliminal</b>	C0 C1	Regarding the measure of reproducing subliminal messages in film and television advertisements:							
		Do you think it will be effective?							
		(Not effective at all)	0	1	2	3	4	5	(Yes, totally effective)
		Do you think it manipulates behaviour?							
		(Not manipulative at all)	0	1	2	3	4	5	(Yes, totally manipulative)
	C2 C3	Regarding the measure of reproducing subliminal messages in film and television advertisements:							
		Do you think it manipulates behaviour?							
		(Not manipulative at all)	0	1	2	3	4	5	(Yes, totally manipulative)
	C4 C5	Regarding the measure of reproducing subliminal messages in film and television advertisements:							
		Do you think it will be effective?							
		(Not effective at all)	0	1	2	3	4	5	(Yes, totally effective)
<b>Co-variable questionnaire (identical in all experimental conditions)</b>									
Nationality	What is your nationality? <input type="radio"/> Spanish <input type="radio"/> Spanish and other ____ <input type="radio"/> Foreign								
Activity	What is your main activity? <input type="radio"/> I work full time <input type="radio"/> Part time work <input type="radio"/> Retiree or pensioner <input type="radio"/> Unemployed (I have worked before) <input type="radio"/> Unemployed (I'm looking for my first job) <input type="radio"/> Student <input type="radio"/> Housework								
Studies	What is your level of completed studies? <input type="radio"/> Without studies (Primary studies without finishing) <input type="radio"/> First Grade (School certificate, EGB 1st stage, more or less 10 years) <input type="radio"/> Second Grade. 1st Cycle (School graduate, or EGB 2nd stage, 1st and 2nd ESO-1st cycle- up to 14 years) <input type="radio"/> Second Grade. 2nd Cycle (FP I and II, Higher Bachelor, BUP, 3rd and 4th ESO (2nd cycle) COU, PREU, 1st and 2nd, up to 18 years) <input type="radio"/> Third Grade. 1st Cycle (Equivalent to Technical Engineer, 3 years, University Schools, Technical Engineers, Technical Architects, Experts,								

	<p>Teaching, ATS, University Graduates, 3-year career, Social graduates, Social workers, etc)</p> <ul style="list-style-type: none"> <li>○ Degree, Degree. 2nd Cycle (University, Higher Graduates, Faculties, Higher Technical Schools, etc.</li> <li>○ Third Degree (Master)</li> <li>○ Third grade (Doctorate)</li> </ul>
Ideology	<p>When talking about politics, the expressions "left" and "right" are commonly used. Where would you place your opinion / political position? (0 = extreme left; 10 = extreme right)</p> <p>○      ○      ○      ○      ○      ○      ○      ○      ○      ○      ○</p> <p>0      1      2      3      4      5      6      7      8      9      10</p> <p>or</p> <ul style="list-style-type: none"> <li>○ I prefer not to answer</li> </ul>
Trust	<p>Do you trust the government and the public administration to develop politics to promote social wellbeing? (1 = Not at all, 5 = Yes, totally)</p> <p>○      ○      ○      ○      ○      ○</p> <p>0      1      2      3      4      5</p>
Support to promote healthy eating	<p>In general, are you in favour of the government intervening to promote healthy eating? (1 = No, not at all in favour; 5 = Yes, very much in favour)</p> <p>○      ○      ○      ○      ○      ○</p> <p>0      1      2      3      4      5</p>
Opinion interference with individual choices	<p>The administration should: (1 = put limits on individual choices to promote individual well-being, 5 = not intervene in their decisions, even if they have negative consequences for their well-being)</p> <p>○      ○      ○      ○      ○      ○</p> <p>0      1      2      3      4      5</p>
Perception of self-diet	<p>In general, do you consider yourself a person who has a healthy diet? (1 = No, very unhealthy; 5 = Yes, very healthy)</p> <p>○      ○      ○      ○      ○      ○</p> <p>0      1      2      3      4      5</p>
Importance healthy diet	<p>Is it important for you to maintain a healthy diet? (1 = No, not at all; 5 = Yes, a lot)</p> <p>○      ○      ○      ○      ○      ○</p> <p>0      1      2      3      4      5</p>
Beliefs sugar effect on personal health	<p>Do you think your health is affected by the consumption of sugary drinks? (1 = No, nothing; 5 Yes, a lot)</p> <p>○      ○      ○      ○      ○      ○</p> <p>0      1      2      3      4      5</p>
Attention to sugar	<p>Usually, do you look at the sugar content of the products you consume? (1 = No, nothing; 5 Yes, a lot)</p> <p>○      ○      ○      ○      ○      ○</p> <p>0      1      2      3      4      5</p>
Consumption of SSBs	<p>How often do you consume sugary drinks?</p> <ul style="list-style-type: none"> <li>○ More than 2 times a day</li> <li>○ Once a day</li> <li>○ 2 to 4 times a day</li> <li>○ Once a week</li> <li>○ Occasionally (less than once a week)</li> </ul>

	<input type="radio"/> Rarely or never
Diet	Are you or have you been on a diet in the last 12 months? <input type="radio"/> Yes <input type="radio"/> No
Perception of need for external help, self	In general, do you consider that in your day-to-day life you would benefit from some kind of help to have a more balanced and healthy diet? (1 = No, nothing; Yes, a lot) <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5
Situationism self	Do you consider that your behaviour is mainly determined... (1 = By my choices and decisions, or 5 = By the context and circumstances that surround me) <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5
Situationism others	Do you consider that the behaviour of the majority of people is mainly determined... (1 = By their choices and decisions, 5 = By the context and circumstances that surround them) <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5
Beliefs obesity	Why do you think people who are overweight and obese? (1 = Because there are too many opportunities to access unhealthy products and difficulties to live a healthy life or 5 = Because people have self-control problems and do not correct their behaviour) <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5
Perception of need for external help, others	In general, do you consider that most people in their day to day life would benefit from some kind of help to have a more balanced and healthy diet? (1 = No, nothing; Yes, a lot) <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5
height	Could you indicate your height? _____cm or <input type="radio"/> I do not know
weight	Could you indicate your weight? _____kg or <input type="radio"/> I do not know

## 1.2. Sample characteristics Study A

**Table 18. Sample characteristics participants Study A**

Factors	Distribution of respondents
Age	Range: 19-79 Median: 45 Mean + SD: 46.69 + 15.69
Biological sex	Male: 49.68% Female: 50.31%
Highest education level	Unfinished education or non: 1.25% School: 2.81% High school (1): 7.81% High school (2): 40.63% College (1): 17.18% College or university degree (2): 18.43% Port-graduate degree (master's): 8.12% Post-graduate degree (PhD): 3.75%
Activity	Full-time job: 46.87% Part-time job: 10.00% Retired: 22.81% Unemployed: 8.43% Unemployed (fist job): 0.62% Student: 7.18% Housework: 4.06%
Region (CCAA)	Andalucía: 14.06% Aragón: 3.75% Principado de Asturias: 3.75% Illes Balears: 1.25% Canarias: 4.06% Cantabria: 0.93% Castilla y León: 10.00% Castilla-la Mancha: 5.62% Catalunya: 16.56% Comunitat Valenciana: 7.5% Extremadura: 1.56% Galicia: 5.31% Madrid: 18.75% Murcia: 2.18% Navarra: 0.31% País Vasco: 4.06% La Rioja: 0.312% Ceuta: 0% Melilla: 0%
Nationality	Spanish: 95.62% Spanish and other: 2.81%

	Others: 1.56%
Ideology	(0=extreme left, 10= extreme right) 0: 1.25% 1: 1.25% 2: 15.00% 3: 15.93% 4: 8.43% 5: 17.50% 6: 8.12% 7: 6.56% 8: 4.68% 9: 1.25% 10: 1.25% Na: 18.75%
Trust	Not at all: 22.81% No: 22.81% Neutral: 34.68% Yes: 15.31% Very much so: 4.37%
Support for government interference to promote healthy eating	Not at all: 3.12% No: 2.50% Neutral: 19.68% Yes: 30.31% Very much so: 44.37%
Should the public administration interference with individual choices to promote individual wellbeing	The 1: 11.56% 2: 18.43% 3: 42.81% 4: 16.25% 5: 10.93%
Self-perception of the quality of their diet	1: 1.25% 2: 2.18% 3: 34.68% 4: 45.93% 5: 15.93%
Importance of maintaining a healthy diet	1: 0% 2: 0.62% 3: 12.81% 4: 39.37% 5: 47.18%
Effect of consumption of SSB on health	1: 21.56% 2: 19.37% 3: 17.81% 4: 21.56% 5: 19.68%
Sugar consciousness	1: 13.12%

	<p>2: 16.56%</p> <p>3: 18.12%</p> <p>4: 27.18%</p> <p>5: 25.00%</p>
Consumption of SSB	<p>More than twice a day: 3.12%</p> <p>Once a day: 6.56%</p> <p>From twice to four times a week: 17.18%</p> <p>Once a week: 15.00%</p> <p>Rarely: 28.12%</p> <p>Never: 30%</p>
Diet on the last 12 months	<p>Yes: 25.31%</p> <p>No: 74.68%</p>
BMI	<p>Range: 16.65 – 62.89</p> <p>Median: 24.61</p> <p>Mean: 25.49</p>
Perceived need for help in making healthier choices for oneself	<p>1: 13.43%</p> <p>2: 10.00%</p> <p>3: 30.93%</p> <p>4: 26.56%</p> <p>5: 19.06%</p>
Perceived need for help in making healthier choices for others	<p>1: 7.18%</p> <p>2: 14.68%</p> <p>3: 36.25%</p> <p>4: 24.37%</p> <p>5: 17.50%</p>
Beliefs regarding the causes of overweight and obesity	<p>1: 23.12%</p> <p>2: 13.75%</p> <p>3: 18.75%</p> <p>4: 21.87%</p> <p>5: 22.50%</p>
Self-perception of dispositionism vs situationism	<p>My behaviour is mainly determined (1 = by my choices and decisions, 5 = by the context and the circumstances that surround me)</p> <p>1: 29.68%</p> <p>2: 20.00%</p> <p>3: 29.37%</p> <p>4: 13.75%</p> <p>5: 7.18%</p>
Others perception of dispositionism vs situationism	<p>The behaviour of most people is mainly determined (1 = by their choices and decisions, 5 = by the context and circumstances around them)</p> <p>1: 6.25%</p> <p>2: 7.5%</p> <p>3: 30.00%</p> <p>4: 33.43%</p> <p>5: 22.81%</p>

### 1.3. Data preparation

Some independent variables had to be transformed to meet the requirements of robustness before proceeding with the statistical analysis. To analyse the relationship between categorical variables with the chi-square test several of the variable's categories were combined, increasing the number of observations in the subgroups and ensuring that none of the categories' expected frequencies were lower than 5%. The recoded variables were also used for the logistic regression.

*Education attainment* was measured in 8 categories. The number of observations in some of the categories was too small. To ensure statistical relevance, the eight original categories were combined into three categories: (1) less than secondary studies, combining the categories: unfinished education or non and primary first and second stage school certificate; (2) secondary studies, including the original category second cycle secondary degree education, (3) university studies, combining two categories of university education of first and second cycle, and master's degree and PhD education graduates.

*Activity* was measured in 7 categories. The category 'unemployed looking for first job' was residual. To ensure statistical relevance, it was combined with the category unemployed, creating a single category for those that do not work but are looking for a job. Activity had an abnormal distribution, 46.87% of the sample is working full-time, and the rest of the categories are less frequent. It is, however, not possible to group categories because it does not make theoretical sense. The variable is excluded from the regression models because it was not significantly related with the 8 dependent variables and the non-employed categories had very few cases.

The variable *Nationality* was measured using three categories (Spanish, Spanish and others and Others) it was discharged for the analysis because only 4% of the sample did not have the Spanish nationality, with such level it failed to make an impact on the models.

*Ideology* was measured initially with a 0 to 10 Likert-type scale. Due to the small number of observations in some categories, I gathered some levels: 0,1,2,3 responses were categorised as 'left', 4, 5, 6 responses were categorised as 'centre', 7,8,9,10 responses were categorised as 'right'. Respondents that choose not to answer were treated as missing values (18.48%)

*Trust in institutions*, *Support for government interference to promote healthy eating* and *Support for government interference with individual choices for individual good*, were collected using 5-point Likert-scales, ranging from no 1 ('trust/strongly disagree') to 5 ('strong trust/strongly agree'). To ensure the robustness of the statistical analysis, the five original categories were recorded in three: 1 ('no-trust in institutions') and ('no support for government interference'), combining 1 and 2 response; ('neutral') including the 3 responses and ('trust in institutions'), ('support for government interference for individual good') and ('support for government interference for social good') combining 4 and 5 responses.

*Perceived healthiness of self-diet* was collected using a 1 ('Very unhealthy') to 5 ('Very healthy') Likert-scale. As expected, not many people declared that their diet was unhealthy. To ensure the robustness of the statistical analysis, the five original categories were recorded in three: perception of unhealthiest or neither healthy or unhealthy (combining 1, 2 and 3 responses), perception of having a healthy diet (including 4 responses), and perception of having a very healthy diet (including 5 responses). The variable was drop for the multilevel regression due to problems of multicollinearity.

The *importance to maintain a healthy diet*, the *beliefs about the impact that the consumption of SSB has on health*, and *degree of attention to sugar in one's diet* were originally collected using 5-point-Likert scales ranging from 1 ('No, at all') to 5 ('Yes, totally'). To ensure the robustness of the statistical analysis, the five original categories were recorded in three with values 1 (combining 1 and 2), 2 (including 3 responses) and 3 (including 4 and 5 responses).

*Consumption of SSBs* was originally collected using 6 categories. To ensure the robustness of the statistical analysis, the two first categories (those that reported having SSBs more than once or once a day) were recoded into a single category 1 ('daily consumption').

*BMI* was calculated by dividing the self-reported body weight (in kg) by the square of the respondent's height (in m<sup>2</sup>).

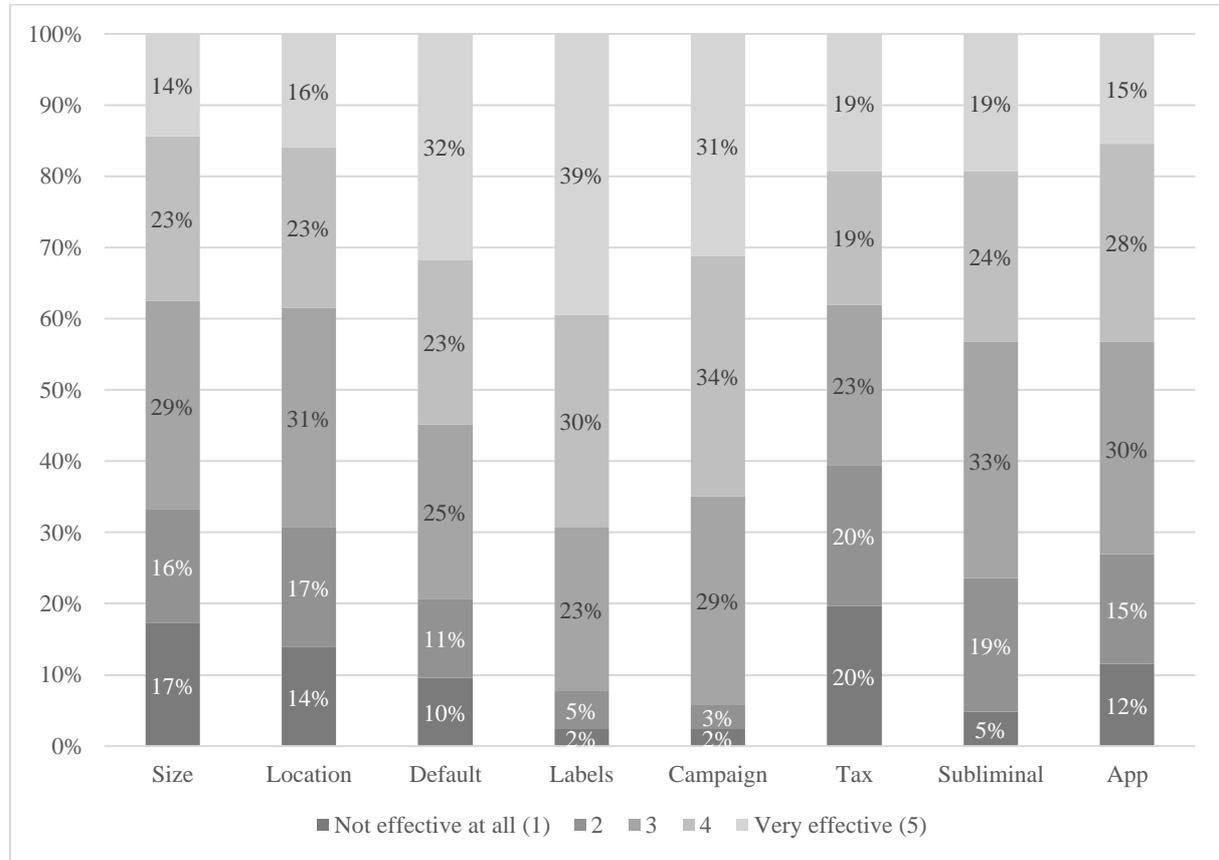
*Perception of ones and others need for external help to maintain a healthy diet* were original collected with 5-point-Likert scales ranging from 1 ('Not at all') to 5 ('Yes totally'). Given the distribution of the data, in both cases, the five original categories were recorded in to three: 1 (combining 1 and 2 responses), 2 (including 3 response) and 3 (combining 4 and 5 response).

*Perceptions of self and other situationism vs dispositionism* were also originally recorded with 5-point-Likert scales, ranging from 1 ('Because of my/their choices and decisions') to 5 ('Because of the context and circumstances that surround me/others'). In the question about oneself, most responses concentrated in the dispositionism range. Given the distribution of the data, the 4 and 5 responses where combined. In contrast, in the question about the tendencies of the others, most responses concentrated in the situationism range. Given the distribution of the data, the 1 and 2 responses where combined.

Finally, the question on beliefs *about the causes of obesity* was also recoded. The original variables collected responses in a 5-point-Likert Scale ranging from 1 ('Because there are too many opportunities to access healthy products and difficulties to lead a healthy life') to 5 ('Because people have self-control problems and do not correct their behaviour. Given the distribution of the data, the 5 original categories were recoded into 3: 1 (combining 1 and 2 responses), 2 (including 3 response) and 3 (combining 4 and 5 response).

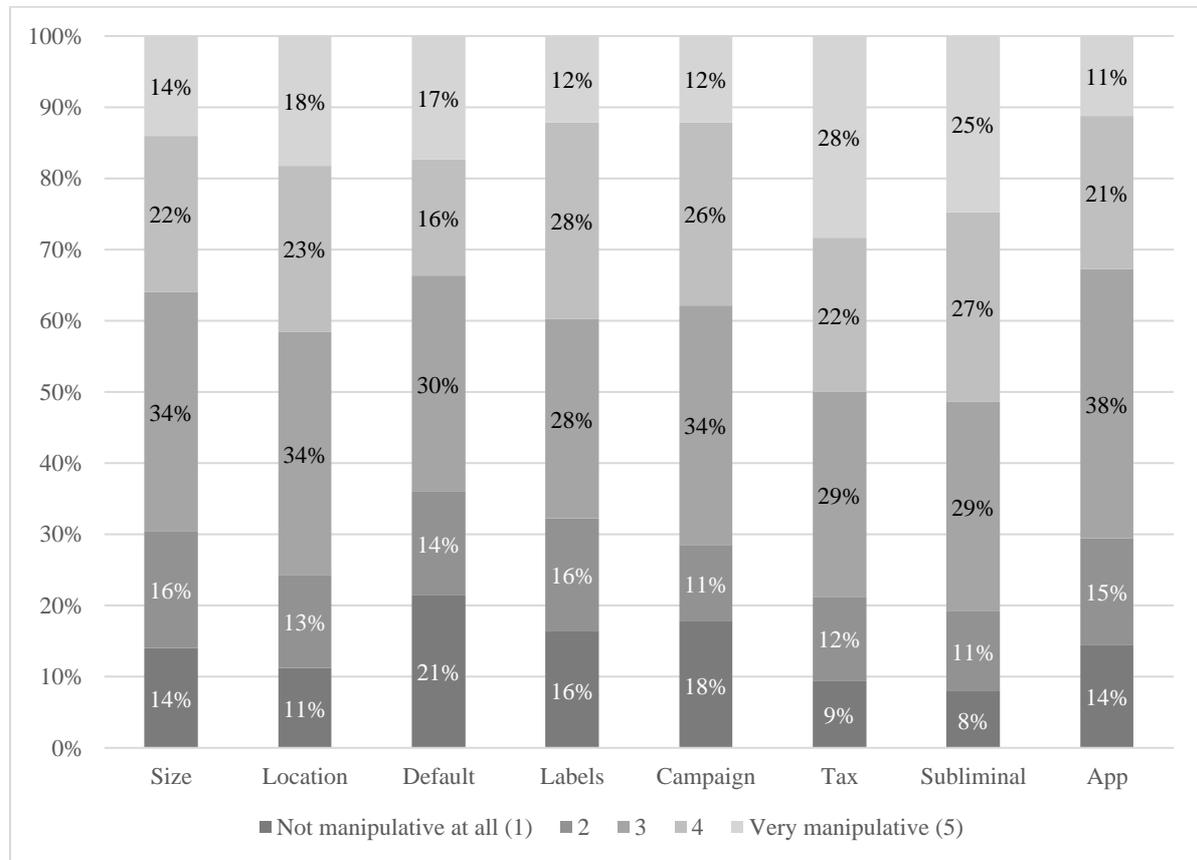
## 1.4. Data on perceived effectiveness and perceived manipulation

**Figure 9. Perceived effectiveness of the eight interventions by response.**



Percentage of responses to the questions ‘Do you think the intervention would be effective?’ by category for each intervention. n= 210, data from conditions: 0, 1, 4 and 5. Automatic nudges: size, location and default. Reflective nudges: labels, campaign and app. No-nudge measures: subliminal messages and tax.

**Figure 10. Perceived manipulation of the eight interventions by response.**



Percentage of responses to the questions ‘Do you think the intervention manipulates behaviour?’ by category for each intervention. n= 214, data from conditions: 0, 1, 2 and 3. Automatic nudges: size, location and default. Reflective nudges: labels, campaign and app. No-nudge measures: subliminal messages and tax.

## 2. Supplementary materials Study B.

### 2.1. Vignettes and survey for Study B.

**Table 19. Contrastive vignettes used in the healthy eating scenario (A)**

	<b>Situationist</b>	<b>Dispositionist</b>
<b>Automatic Nudge</b>	<p>To promote healthy eating, the administration has decided to regulate the distribution of unhealthy products in supermarkets so that they are placed in are less visible and difficult to access areas.</p> <p>The supermarket where Martin buys has adopted the measure. Usually, Martin is a person who does not always pay attention to his surroundings. The context and options available to him tend to affect his decisions and he does not mind getting carried away.</p> <p>The measure has had a strong influence on his behaviour and has affected his decisions in a totally unconscious way, beyond his perception and control.</p> <p>As a result of the intervention, Martin has improved his diet. However, Martin has not realized that changing the location of the products has influenced his decisions. Martin does not know that he has chosen healthier products due to the effect of external influence and not by personal choice.</p>	<p>To promote healthy eating, the administration has decided to regulate the distribution of unhealthy products in supermarkets so that they are placed in are less visible and difficult to access areas.</p> <p>The supermarket where Martin buys has adopted the measure. Usually, Martin is a person who chooses according to his preferences, who likes to be in control and make his own decisions.</p> <p>Despite its character, the measure has had a strong influence on their behaviour and has affected his decisions in a totally unconscious way, beyond their perception and control.</p> <p>As a result of the intervention, Martin has improved his diet. However, Martin has not realized that changing the location of the products has influenced his decisions. On this occasion, Martin does not know that he has chosen healthier products due to the effect of external influence and not by personal choice.</p>
<b>Reflective Nudge</b>	<p>To promote healthy eating, the administration has decided to introduce informational traffic-light like labels on the packaging of different products and marking high-sugar beverages in red to highlight that they are unhealthy.</p> <p>The supermarket where Martin buys has adopted the measure. Usually, Martin is a person who does not always pay attention to his surroundings. The context and options available to him tend to affect his decisions and he does not mind getting carried away</p> <p>The measure has had a strong influence on his behaviour and has made him reflect. It has affected his decisions in a totally conscious way.</p> <p>As a result of the intervention, Martin has improved his diet. In addition, Martin has realized that labels have influenced his decisions and led him to choose healthier options.</p>	<p>To promote healthy eating, the administration has decided to introduce informational traffic-light like labels on the packaging of different products and marking high-sugar beverages in red to highlight that they are unhealthy.</p> <p>The supermarket where Martin buys has adopted the measure. Usually, Martin is a person who chooses according to his preferences, who likes to be in control and make his own decisions.</p> <p>The measure has had a strong influence on his behaviour and has made him reflect. It has affected his decisions in a totally conscious way.</p> <p>As a result of the intervention, Martin has improved his diet. In addition, Martin has realized that labels have influenced his decisions and led him to choose healthier options.</p>

**Table 20. Contrastive vignettes used in the organ donation scenario (B)**

	<b>Situationist</b>	<b>Dispositionist</b>
<b>Automatic Nudge</b>	<p>To increase the number of organ donors and to meet the demands of patients who need a transplant, the administration has decided to introduce a change in the consent form and preselect the option "be a donor".</p> <p>This measure has had a strong influence on Martin's decision, who, when filling in the form, has not modified the option.</p> <p>Normally, Martin is a person who does not always decide carefully. The context and options available to him affect his decisions and he does not mind getting carried away.</p> <p>The measure has affected his behaviour in a totally unconscious way, beyond his perception and control. In other words, Martin has not realized that the default option has influenced his final decision to register as an organ donor.</p>	<p>To increase the number of organ donors and to meet the demands of patients who need a transplant, the administration has decided to introduce a change in the consent form and preselect the option "be a donor".</p> <p>This measure has had a strong influence on Martin's decision, who, by filling in the form, has not modified the option. Normally, Martin is a person who chooses according to his preferences, who likes to be in control and make his own decisions.</p> <p>Despite his character, the measure has affected his behaviour, totally unconsciously, beyond his perception and control. On this occasion, Martin has not realized that the default option has influenced his final decision to register as an organ donor.</p>
<b>Reflective Nudge</b>	<p>To increase the number of organ donors and to meet the demands of patients who need a transplant, the administration has decided to introduce a change in the consent form and preselect the option "be a donor".</p> <p>This measure has had a strong influence on Martin's decision, who, when filling in the form, has not modified the option.</p> <p>Normally, Martin is a person who does not always decide carefully. The context and options available to him affect his decisions and he does not mind getting carried away.</p> <p>The measure has consciously affected his behaviour and made him reflect. In other words, Martin has realized that the default option has influenced his final decision to register as an organ donor.</p>	<p>To increase the number of organ donors and to meet the demands of patients who need a transplant, the administration has decided to introduce a change in the consent form and preselect the option "be a donor".</p> <p>This measure has influenced Martin's decision, who, when filling in the form, has not modified the option. Normally, Martin is a person who chooses according to his preferences, who likes to be in control and make his own decisions.</p> <p>In this case, the measure has consciously affected his behaviour and made him reflect. In other words, Martin has realized that the default option has influenced his final decision to register as an organ donor.</p>

**Table 21. Experimental sequence and questions Study B**

Screen and function	Text
1 Initial instructions	Next, you will see a short text about a hypothetical scenario. Please read it carefully.
2 Vignette + first two questions	<p>TEXT vignette</p> <hr/> <p>Considering what you know about Martin, what do you feel about influencing his decisions in this way?</p> <p>Use the slider to mark your answer. Indicate if it is closer to "Very bad" or "Very good".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between; align-items: center;"> <span data-bbox="549 741 624 808">Very bad</span>  <span data-bbox="1155 741 1230 808">Very good</span> </div> <p>Can you briefly indicate what has motivated your answer in the previous question?</p> <div style="border: 1px solid black; height: 60px; width: 100%;"></div>
3 Manipulation checks	<p>How is Martin described? :</p> <ul style="list-style-type: none"> <li><input type="radio"/> Someone who is affected by context and does not mind letting go.</li> <li><input type="radio"/> Someone who likes to be in control and make their own decisions.</li> </ul> <p>How is the influence of the measure on Martin's behaviour described?</p> <ul style="list-style-type: none"> <li><input type="radio"/> Conscious influence</li> <li><input type="radio"/> Unconscious influence</li> </ul>
4 Questions on intrusiveness	<p>Do you think that the measure manipulates Martin's decisions?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not at all" or "Yes, totally".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between; align-items: center;"> <span data-bbox="549 1503 624 1570">Not at all</span>  <span data-bbox="1155 1503 1230 1570">Yes, totally</span> </div> <p>Do you think that the measure restricts Martin's freedom of choice?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not at all" or "Yes, totally".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between; align-items: center;"> <span data-bbox="549 1738 624 1805">Not at all</span>  <span data-bbox="1155 1738 1230 1805">Yes, totally</span> </div> <p>Do you think Martin would accept the intervention?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not at all" or "Yes, totally"</p>

	<div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between;"> <span>Not at all</span>  <span>Yes, totally</span> </div> <p>Do you think it is ethical to influence Martin's behaviour in this way?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not ethic at all" or "Yes, totally".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between;"> <span>Not ethic at all</span>  <span>Yes, totally</span> </div> <p>Do you consider that the effect of the measure on Martín's behaviour is positive?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not at all" or "Yes, totally".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between;"> <span>Not at all</span>  <span>Yes, totally</span> </div>
5 Questions on acceptability	<p>Would you oppose the implementation of the measure?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not at all" or "Yes, totally".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between;"> <span>Not at all</span>  <span>Yes, totally</span> </div>
	<p>Do you think the measure is acceptable?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not at all" or "Yes, totally".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between;"> <span>Not at all</span>  <span>Yes, totally</span> </div>
	<p>Would you be in favour of applying this measure?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not at all" or "Yes, totally".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between;"> <span>Not at all</span>  <span>Yes, totally</span> </div>
6 Question on effectivity	<p>Do you think that the measure described will be effective?</p> <p>Use the slider to mark your answer. Indicated if it is closer to "Not effective at all" or "Yes, very effective".</p> <div style="border: 1px solid black; padding: 5px; display: flex; justify-content: space-between;"> <span>Not effective at all</span>  <span>Yes, very effective</span> </div>
<b>Co – variables questionnaire</b>	
7 Nationality	<p>What is your nationality?</p> <ul style="list-style-type: none"> <li><input type="radio"/> Spanish</li> <li><input type="radio"/> Spanish and other_____</li> <li><input type="radio"/> Foreign</li> </ul>
8 Activity	<p>What is your main activity?</p> <ul style="list-style-type: none"> <li><input type="radio"/> I work full time</li> <li><input type="radio"/> Part-time work</li> </ul>

	<ul style="list-style-type: none"> <li>○ Retiree or pensioner</li> <li>○ Unemployed (I have worked before)</li> <li>○ Unemployed (I'm looking for my first job)</li> <li>○ Student</li> <li>○ Housework</li> </ul>
9 Studies	<p>What is your level of completed studies?</p> <ul style="list-style-type: none"> <li>○ Without studies (Primary studies without finishing)</li> <li>○ First Grade (School certificate, EGB 1st stage, more or less 10 years)</li> <li>○ Second Grade. 1st Cycle (School graduate, or EGB 2nd stage, 1st and 2nd ESO-1st cycle- up to 14 years)</li> <li>○ Second Grade. 2nd Cycle (FP I and II, Higher Bachelor, BUP, 3rd and 4th ESO (2nd cycle) COU, PREU, 1st and 2nd Bachillerato, up to 18 years)</li> <li>○ Third Grade. 1st Cycle (Equivalent to Technical Engineer, 3 years, University Schools, Technical Engineers, Technical Architects, Experts, Teaching, ATS, University Graduates, 3-year career, Social graduates, Social workers, etc)</li> <li>○ Degree, Degree. 2nd Cycle (University, Higher Graduates, Faculties, Higher Technical Schools, etc.</li> <li>○ Third Degree (Master)</li> <li>○ Third grade (Doctorate)</li> </ul>
10 Ideology	<p>When talking about politics, the expressions "left" and "right" are commonly used. Where would you place your opinion / political position? (0 = extreme left; 10 = extreme right)</p> <p>○   ○   ○   ○   ○   ○   ○   ○   ○   ○   ○ 0   1   2   3   4   5   6   7   8   9   10</p> <p>or</p> <p>○ I prefer not to answer</p>
11 Trust	<p>Do you trust the government and the public administration to develop politics to promote social wellbeing? (1 = Not at all, 5 = Yes, totally)</p> <p>○   ○   ○   ○   ○ 0   1   2   3   4   5</p>
10 Situationism self	<p>Do you consider that your behaviour is mainly determined...? (1 = By my choices and decisions, or 5 = By the context and circumstances that surround me)</p> <p>○   ○   ○   ○   ○   ○ 0   1   2   3   4   5</p>
11 Situationism others	<p>Do you consider that the behaviour of the majority of people is mainly determined...? (1 = By their choices and decisions, 5 = By the context and circumstances that surround them)</p> <p>○   ○   ○   ○   ○   ○ 0   1   2   3   4   5</p>
12 Beliefs obesity or Organ donor	<p>Why do you think people who are overweight and obese? (1 = Because there are too many opportunities to access unhealthy products and difficulties to live a healthy life or 5 = Because people have self-control problems and do not correct their behaviour)</p> <p>○   ○   ○   ○   ○   ○ 0   1   2   3   4   5</p> <p>Are you an organ donor?</p> <ul style="list-style-type: none"> <li>○ Yes</li> <li>○ No</li> </ul>

## 2.2. Sample characteristics Study B.

**Table 22. Sample characteristics participants Study B**

<b>Factors</b>	<b>Distribution of respondents</b>
Age	Range: 19-83 Median: 38 Mean: 40.54 + SD: 13.54
Biological sex	Male: 49.12% Female: 50.87%
Education level	Primary studies or least:4.82% Secondary studies: 33.77% University studies: 46.05% Post-graduate studies: 15.35%
Activity	Full-time job: 60% Part-time job: 12.28 % Retired: 7.45% Unemployed:7.89 % Unemployed (fist job):0.87 % Student: 8.33% Housework: 3.07%
Ideology	(0=extreme left, 10= extreme right) 0: 3.94% 1: 2.63% 2: 12.28% 3: 19.73% 4: 11.40% 5:19.29% 6: 3.50% 7: 8.33% 8: 4.38% 9:1.75 % 10: 2.19% NA: 10.52%
Trust	Not at all: 21.05% No: 28.07% Neutral: 31.14% Yes: 15.35% Very much so: 4.38%
Self-perception of dispositionism vs situationism	My behaviour is mainly determined (1=by my choices and decisions, 5 =by the context and the circumstances that surround me) 1:17.89% 2: 23.68% 3: 39.91% 4: 14.47% 5: 3.94%

<p>Perception of others dispositionism vs situationism</p>	<p>The behaviour of most people is mainly determined (1=by their choices and decisions, 5 =by the context and circumstances around them)</p> <p>1: 2.63%</p> <p>2: 6.14%</p> <p>3: 35.08%</p> <p>4: 39.91%</p> <p>5: 16.22%</p>
<p>Beliefs regarding the causes of overweight and obesity</p>	<p>Why do you think people are overweight and obese (1=Because there are too many opportunities to access unhealthy products and difficulties in leading a healthy life, 5=Because people have self-control problems and do not correct their behaviour)?</p> <p>1: 22.52%</p> <p>2: 17.11%</p> <p>3: 26.12%</p> <p>4: 39.91%</p> <p>5: 16.21%</p>
<p>Organ donors</p>	<p>Yes: 30.76</p> <p>No: 69.23</p>

### 2.3. Pairwise comparisons Study B.

Scenario A. Healthy eating					Scenario B. Organ donation				
<b>How do you feel?</b>									
	diff	lwr	upr	p		diff	lwr	upr	p adj
AS-AD	9,12	-11,28	29,53	0,649	AS-AD	-8,04	-29,31	13,23	0,758
RD-AD	25,15	2,22	48,08	0,026	RD-AD	-10,58	-34,04	12,88	0,643
RS-AD	10,18	-14,35	34,71	0,701	RS-AD	1,58	-27,49	30,65	0,999
RD-AS	16,03	-3,22	35,28	0,137	RD-AS	-2,55	-23,82	18,73	0,989
RS-AS	1,06	-20,07	22,19	0,999	RS-AS	9,61	-17,72	36,95	0,796
RS-RD	-14,97	-38,55	8,61	0,352	RS-RD	12,16	-16,91	41,23	0,696
<b>How manipulative?</b>									
	diff	lwr	upr	p		diff	lwr	upr	p adj
AS-AD	1,27	-14,99	17,52	0,997	AS-AD	-4,57	-20,68	11,54	0,881
RD-AD	-36,31	-54,59	-18,04	0,000	RD-AD	-19,17	-36,94	-1,40	0,029
RS-AD	-13,06	-32,60	6,49	0,307	RS-AD	-18,85	-40,87	3,17	0,121
RD-AS	-37,58	-52,92	-22,24	0,000	RD-AS	-14,60	-30,71	1,51	0,090
RS-AS	-14,32	-31,16	2,51	0,124	RS-AS	-14,27	-34,98	6,43	0,280
RS-RD	23,26	4,47	42,04	0,009	RS-RD	0,32	-21,70	22,34	1,000
<b>Limits freedom of choice?</b>									
	diff	lwr	upr	p adj		diff	lwr	upr	p adj
AS-AD	-3,44	-22,05	15,16	0,963	AS-AD	-7,38	-26,31	11,55	0,740
RD-AD	-31,10	-52,01	-10,19	0,001	RD-AD	-24,36	-45,24	-3,49	0,015
RS-AD	-30,79	-53,16	-8,42	0,003	RS-AD	-28,28	-54,16	-2,41	0,026
RD-AS	-27,66	-45,21	-10,10	0,000	RD-AS	-16,98	-35,91	1,95	0,095
RS-AS	-27,35	-46,61	-8,08	0,002	RS-AS	-20,90	-45,23	3,42	0,119
RS-RD	0,31	-21,19	21,81	1,000	RS-RD	-3,92	-29,79	21,95	0,979
<b>Would Martin accept?</b>									
	diff	lwr	upr	p adj		diff	lwr	upr	p adj
AS-AD	11,64	-3,82	27,10	0,208	AS-AD	8,07	-7,11	23,26	0,511
RD-AD	14,13	-3,24	31,51	0,152	RD-AD	8,09	-8,66	24,84	0,590
RS-AD	32,20	13,61	50,79	0,000	RS-AD	32,76	12,01	53,51	0,000
RD-AS	2,49	-12,10	17,08	0,970	RD-AS	0,02	-15,17	15,20	1,000
RS-AS	20,56	4,55	36,57	0,006	RS-AS	24,69	5,18	44,21	0,007
RS-RD	18,07	0,20	35,93	0,046	RS-RD	24,67	3,92	45,42	0,013

Scenario A. Healthy eating					Scenario B. Organ donation				
How ethical?									
	diff	lwr	upr	p adj		diff	lwr	upr	p adj
AS-AD	4,93	-12,15	22,01	0,875	AS-AD	-4,65	-20,38	11,08	0,867
RD-AD	21,47	2,28	40,67	0,022	RD-AD	16,86	-0,49	34,20	0,060
RS-AD	28,26	7,72	48,79	0,003	RS-AD	18,22	-3,28	39,71	0,127
RD-AS	16,54	0,43	32,66	0,042	RD-AS	21,51	5,78	37,23	0,003
RS-AS	23,32	5,64	41,01	0,005	RS-AS	22,87	2,65	43,08	0,020
RS-RD	6,78	-12,96	26,52	0,807	RS-RD	1,36	-20,13	22,86	0,998
Would the effect be positive?									
	diff	lwr	upr	p adj		diff	lwr	upr	p adj
AS-AD	-6,35	-20,90	8,20	0,666	AS-AD	2,76	-13,66	19,17	0,972
RD-AD	-3,18	-19,54	13,17	0,957	RD-AD	20,65	2,55	38,75	0,019
RS-AD	10,67	-6,83	28,16	0,388	RS-AD	28,01	5,57	50,44	0,008
RD-AS	3,16	-10,56	16,89	0,931	RD-AS	17,89	1,48	34,31	0,027
RS-AS	17,01	1,95	32,08	0,020	RS-AS	25,25	4,15	46,34	0,012
RS-RD	13,85	-2,96	30,67	0,144	RS-RD	7,35	-15,08	29,79	0,828
Would you support?									
	diff	lwr	upr	p adj		diff	lwr	upr	p adj
AS-AD	-3,63	-22,01	14,74	0,955	AS-AD	-6,47	-24,30	11,36	0,780
RD-AD	15,39	-5,26	36,04	0,216	RD-AD	10,06	-9,61	29,72	0,544
RS-AD	19,52	-2,57	41,61	0,103	RS-AD	17,93	-6,44	42,29	0,226
RD-AS	19,02	1,68	36,36	0,026	RD-AS	16,52	-1,31	34,35	0,080
RS-AS	23,15	4,12	42,18	0,010	RS-AS	24,39	1,48	47,31	0,032
RS-RD	4,13	-17,11	25,37	0,957	RS-RD	7,87	-16,50	32,24	0,834

