An offer they can refuse? Nudging against advertising

by

Thomas Giacoletto

Submitted to

Central European University

Department of Political Science

In partial fulfilment of the requirements for the degree of

Master of Arts

Supervisor: Dr. Andrés Moles

Budapest, Hungary

2018

Acknowledgements

I would like to thank my supervisor, Dr. Andrés Moles, for guiding me throughout the process. I am grateful for all the insightful discussions, as well as the time you spend commenting on my drafts. Your suggestions really improved both my thought and this thesis, and I am very thankful for this.

I also wish to thank Christophe Heintz and Anand Murugesan for teaching a course allying cognitive science and policy making, and for introducing me to many of the topics I explored this year, and specifically in this thesis. You both created a great learning environment during the course, and I do hope that you continue to teach it (together).

I am also grateful to Zsuzsanna Toth, from CEU's Center for Academic Writing, who has read many earlier drafts of several parts of this thesis. Your comments really improved my writing, and I hope that you will enjoy the final version.

And finally, I want to thank Oliver Antczak, Julia Aurik, Marleen Bornat, Sarah Gharib, James Herbert, Gideon de Jong, Serenay Kaykaç, and Nils Oellerich for taking the time (/bothering) to read my drafts. Your comments really improved this thesis, I hope that it was not boring, and I cannot wait to review yours.

Abstract

In this thesis, I investigate whether governments can legitimately use nudges to counter manipulative advertisements. I argue that they can implement counter-manipulative nudges when they are easily resistible, and fulfil both transparency and publicity requirements. Governments can apply nudges since they do not require more justification than traditional policies, are not manipulative, and respect individuals' autonomy. After explaining my argument and presenting two ways in which nudges can be used to counter manipulative advertisement, I defend my position against two objections. I conclude that counter-manipulation in itself is justified because it prevents harm from others, and that nudging is a better option to protect consumers than banning advertisement.

Keywords: public policy, nudges, advertisement, manipulation.

Contents

Introduction

In 2016, in the United Kingdom, private companies spent 27.5 times more on advertisements for junk food than the public services spends on public awareness campaigns, respectively 143 and 5.2 million pounds (O'Dowd 2017). Before proposing a ban on advertisements for junk foods in Transport for London tubes and busses, London mayor Sadiq Khan described advertisement as an "influence and pressure that [is] put on children and families to make unhealthy choices" (BBC 2018). This unhealthy influence and pressure is indeed felt daily by a majority in the UK, and contributes to the 5 billion pounds the National Health Services spends every year in obesity related treatments alone (O'Dowd 2017). While banning advertisements is an option, it is often argued to not have the desired effects (Bruce, Murthi and Rao 2017; Hallsworth 2017; Rayner and Lang 2011; Thaler 2010).

An alternative to using bans is to rely on insights from the behavioural sciences in order to reduce consumers' fast food consumption (Hallsworth 2017; Roberto and Kawachi 2015). In a study testing the effects of labels made compulsory by the United States' Nutrition Labelling and Education Act, Mathios (2000) found that the disclosure of fat levels decreased sales for the fattiest products by 4 to 5.7%, depending on the supermarket observed (667). This means that communicating information to customers in an intelligible way—thanks to visible labels instead of nutrition tables at the back of products—effectively contributes to a reduction of fat consumption. If regular, such a reduction can help reduce obesity and overweight problems, as most cases of overweight are not necessarily caused by large calories excess, but rather by small yet repeated excesses (Hallsworth 2017).

What makes such an intervention successful is the fact that "small and apparently insignificant details can have major impacts on people's behaviour" (Thaler and Sunstein 2008, 3). Thanks to developments in the fields of experimental economics and cognitive psychology, we know that by modifying the "context" surrounding a choice (Allais 1953;

Johnson and Goldstein 2003; Kahneman 2011; Tversky and Kahneman 1981, 1983), it is possible to influence individuals' decisions one way or another (Thaler and Sunstein 2008, 236). Interventions to modify the context surrounding a choice are called choice architecture (CA) modifications, and individuals or organisations modifying CAs are referred to as choice architects (or architects).

Although CA modifications and the use of cognitive biases for public policy have received a lot of positive attention during the last decade, these tools were already used in advertising, although not necessarily formalised for a long time. Product advertisement essentially consists in attempting to change the way consumers see their alternatives when facing a purchasing choice (Packard 1957). While not all forms of advertisement rely on cognitive biases, they all modify consumer's CAs in order to sell more products. Whether it is using emotional appeal (Zhang et al. 2014), simple framing effects (Smith 1996), or priming (Wänke 2016), the goal is to change the vision customers have of certain products (or sometimes to create one) in order to increase sales. In this case, the architect (the advertiser) is modifying consumers' CAs with the goal to sell more products.

The use of behavioural insights and CA modifications is relatively new in policy making (Benartzi et al. 2017; Chetty 2015; Congdon, Kling and Mullainathan 2011; Economist 2017; Haushofer and Fehr 2014; Loewenstein et al. 2015; Thaler and Sunstein 2008). Nudges, which are the most publicised type of behaviourally informed policy intervention, have been named and popularised by Thaler and Sunstein (2008), and have later been endorsed by both academics and policy makers around the world (Benartzi et al. 2017; Economist 2017), although they have not been universally acclaimed (Gigerenzer 2015; Grüne-Yanoff and Hertwig 2015; Hertwig and Grüne-Yanoff 2017). Nudges use CA modifications in order to push (or nudge) individuals towards one choice rather than the one they would presumably have chosen had the nudge not been implemented (Saghai 2013; Thaler and Sunstein 2008). Proponents of their use as public policies claim that they can increase people's welfare

and help them make better choices for themselves while preserving their choice-sets and autonomy (Blumenthal-Barby 2013; Guala and Mittone 2015; Sunstein 2015b; Thaler and Sunstein 2008). Nudges have already been used and are argued to be more effective and cost efficient than traditional (non-behavioural) policies such as incentives or educational programmes (Benartzi et al. 2017; Chetty 2015; Johnson and Goldstein 2003; Madrian and Shea 2001).

Both advertisements and nudges use the same cognitive biases, and rely on CA modifications in order to attain their goal—selling more products for the former, and improving people's welfare for the latter. Advertisements, however, are usually seen as manipulative (Treise et al. 1994; Wilkinson 2013), while nudges are not and seem to instead be welcomed by individuals in the United States of America, Europe, and Asia (Reisch and Sunstein 2016; Sunstein 2015b, 2017a). This means that the reason advertisements are labelled as manipulative is not their reliance on cognitive biases and CA modifications. This difference in conception opposes advertisements and nudges in the sense that the former manipulates and reduces individuals' autonomy, while the latter ensures that they make the best choices for themselves. Beyond the manipulative aspect, there is a second tension between the two concepts, which is that they are likely to clash when applied to the same item.

Let us consider a private company advertising their high-fat, high-sugar foods, and a government using nudges to fight obesity and improve public health. The company will advertise their product hoping to maximise their sales, while the government will attempt to nudge citizens towards a healthier lifestyle—so practically speaking, into not buying the company's products. In this case, consumers' CAs have already been modified by advertisements and would be modified for a second time by the government nudge, in a sense to counter the manipulation induced by the advertisements. The "influence and pressure" in favour of unhealthy foods that Khan described has to be counterbalanced by the CA modifications induced by the government (BBC 2018). The goal for the government

would not be to restore consumers' CAs as if they had been unaltered, as it is not possible to have a neutral CA (Sunstein 2015b; Thaler and Sunstein 2008), but rather to alter them further to prevent the attempted manipulated from being effective.

While most theoretical research focuses on defining whether nudges should also be considered as manipulative, there is little focus on the use of nudges as a counter-manipulation mechanism. There is, furthermore, little consideration for the fact that nudges and advertisements often have conflicting goals in terms of the behaviour they attempt to achieve. In order to consider the possibility of governments nudging as a counter-manipulation mechanism, this thesis is concerned with whether they can legitimately use nudges to prevent citizens from being manipulated by advertisements.

I argue that nudges can indeed be legitimately used to counter manipulation from advertisement. For the purpose of this thesis, I define nudges as (i) human-designed, and intentional, choice architecture modifications which (ii) rely on individuals' cognitive biases to make them take a decision over another, and are (iii) easily resistible. With this definition in mind, I advance that nudges are not different from traditional policies used today in terms of the justification governments need to implement them. There are, however, differences in the way nudges and traditional policies achieve compliance. Despite not requiring further justification, implementing nudges, however, requires a strong focus on their transparency and publicity. Qualifying nudges as either visible or transparent, I argue that nudges which are not fully transparent in themselves—and therefore fully public—are only considered visible, and ought to be made public to be legitimate.

In the first chapter, I review previous literature relevant to the topic to understand the foundations of the nudge approach, as well as how it can be used for counter-manipulation. I first review the literature in experimental economics and psychology which led to the creation of nudges as a policy tool. I then examine arguments for and against nudges, provide a few examples, and present empirical research on their public acceptance. Lastly,

I explore the connection between advertisement and CA modifications, as well as the one between manipulation and behavioural sciences.

In the second chapter, I lay out my argument. I start by setting boundaries for the analysis, before considering the justification needed for the implementation of nudges. For this, I explore whether nudges differ from traditional policies, and how they can counter manipulation by advertisement. I then define nudges depending on whether they are aligned with individuals' preferences or not, and their level of transparency. I conclude by stipulating transparency and publicity requirements for nudges to be used by governments as a counter-manipulation mechanism.

The third chapter addresses two possible objections to counter-manipulation in general, and the use of nudges to this end. I give arguments in favour of counter-manipulation and further argue that nudging yields better outcomes than the alternative: banning advertisement. The last chapter concludes, and indicates other areas in which the use of nudges could be useful and needs to be assessed.

Chapter 1: Theoretical background

The question of whether governments should use nudges as a counter-manipulation measure to prevent citizens from being manipulated by advertisements requires diving into several streams of literature. It is indeed important to have an understanding of human behaviour, as well as of the tools which are being used to direct people towards specific choices, be it for their own good or against it. It is also relevant to have in mind considerations about the concept of manipulation and how it applies to the use of behavioural sciences in altering behaviour. In this chapter, I present the literature which led to the current understanding of human behaviour before presenting two tools informed by this understanding: nudges and advertisement. After introducing the two, I outline literature related to the notions of manipulation and autonomy.

1.1 A better understanding of human behaviour

In the last 50 years, we have moved towards a more complex understanding of human behaviour. In contrast to the rational choice model which has long been the baseline, we now know that we 'misbehave' in the sense that our actions are not always rational. Studying diversions from pure rationality has led to a more accurate understanding of human decisions, and to more accurate predictions, as the outcome of these divergences are predictable. An understanding of the change from the rational choice model to one that integrates behavioural insights is important, and this section reviews the literature which drove this change.

1.1.1 Misbehaving

In an exchange with economist Robert Barro, Richard Thaler stated that the difference "between [their] models was that he [Barro] assumed that the agents in his model were as smart as he was, and I [Thaler] assumed they were as dumb as I am" (Thaler 2015, 222–3).[1] Although intuitive in understanding one of the main differences between rationalist and behavioural economics, Thaler's comment relies on many years of research which allowed him and others to understand how 'dumb' we are. In this section, I review important parts of the evolution which drove the inclusion of behavioural insights within the field of economics.

Economists are often criticised for considering humans as rational decision makers who always rely on cost analysis calculations to flawlessly make choices which are in their best interest (Friedman 1966). However, experimental research influenced by psychology challenged the common assumption of rationality to replace it with a more complete understanding of human behaviour accounting for cognitive biases which often sway human decisions away from rational ones. In 1953, economist Maurice Allais published an article exposing what would later be called the 'Allais paradox' (Allais 1953). He discovered that people do not identify the expected utility of the choices they face in the way predicted by expected utility theory, which assumes rational individuals.

Allais' findings that expectations about human behaviour set by the rational model do not accurately represent human reasoning and cannot accurately predict behaviour fit within the framework of bounded rationality introduced by Herbert Simon (Simon 1955). Initially focusing on decision-making within administrations (Simon 1947), Simon advocated the idea that human decisions were only rational as much as they could given the information and time available, as well as individuals' cognitive limitations (Simon 1955,

[1] Thaler also reports that Barro agreed with the statement.

1985). Almost two decades after Allais' initial contribution, psychologists Amos Tversky and Daniel Kahneman continued research into cognitive biases which lead to consistent deviations from rational decision-making (Kahneman 2003, 2011; Kahneman and Tversky 1979, 1996; Tversky 1969; Tversky and Kahneman 1973, 1974, 1981, 1983).

The work by Tversky and Kahneman introduced and shed light on many heuristics and biases which lead individuals to deviate from what the rational model expects of them (Kahneman and Tversky 1996). Both their research and findings targeted several aspects of human behaviour, and have greatly affected disciplines such as economics and psychology, but also have relevant implications for many other fields including medical and legal professions.

In several publications, they focus on the concept of statistical literacy, and observe that people make systematic mistakes when asked to evaluate probabilities and frequencies (Tversky and Kahneman 1973). The systematic nature of these mistakes was especially relevant for them and they were able to identify heuristics which are used instead of rational thinking. In the case of probabilities and frequencies, they later theorised about three heuristics (representatives, availability and anchoring) people tend to use (Tversky and Kahneman 1974). They further coined the conjunction fallacy, which became known as the 'Linda problem,' and occurs when people over-estimate the probability of two events happening in conjunction as opposed to one of them happening individually (Tversky and Kahneman 1983).[2]

Also important for advertisements and nudges are the topics of framing and prospect theory. Framing is straightforward as a process and has been widely studied by many discip-

[2]They gave respondents a back story about Linda: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations." And asked which one of the two following statements were more probable: 'Linda is a bank teller' or 'Linda is a bank teller and is active in the feminist movement.'
Most respondents responded that the latter one was more likely, using the representativeness heuristics instead of calculating probabilities (Tversky and Kahneman 1983).

lines. It also holds a prominent place in communication research. Thanks to experiments about monetary outcomes and human lives, Tversky and Kahneman (1981) show that the same problem framed in different ways leads to preference reversal. In what became a famous example of framing effects, they presented two groups of respondents with the same problem and the exact same outcomes framed as either saving, or losing lives. The problem describes a disease outbreak which is expected to kill 600 people, and presents two programmes in order to reduce the number of casualties. While the two versions of the problem are "effectively identical," the programmes proposed to the two groups are framed differently (Tversky and Kahneman 1981, 453). The programmes proposed to the first group emphasised the number of lives saved, while the ones proposed to the second group emphasised the number of lives lost.

While the programmes lead to the exact same outcomes in both versions of the problem and respondents should have choose all solutions equally, they find that this is not the case, and framing, by emphasising either gains or losses, led to preference reversal. Indeed, while the less risky option was chosen by 72% of respondents when the framing emphasised gains, it was only chosen by 22% when the framing focused on losses. This finding has drastic implications for the rational choice model since it means that people do not only make choices based on their outcomes, but also on the way the alternatives are framed (Tversky 1969; Tversky and Kahneman 1981). Findings about framing led Kahneman and Tversky to create their own theory of economic behaviour which they called 'prospect theory' (1979). Instead of modelling optimal decisions as the rational choice theory does, prospect theory explains people's choices relative to the values of losses and gains attached to available choices, and not on the outcomes of these choices (Kahneman and Tversky 1979). This means that people evaluate their choices for what they are and how they are framed instead of their outcomes.

1.1.2 Predictably misbehaving

The most important finding that came from Tversky and Kahneman's research is not merely that human behaviour is not rational, but rather that it is *Predictably Irrational* (Ariely 2008; Kahneman 2011). Much of the literature in behavioural and experimental economics has focused on understanding these heuristics and shortcuts which make diversions from the rational choice model predictable. This allows for a better understanding of human behaviour, and the possibility to accurately predict it (Thaler 1980). Predicting behaviour is the reason why we look for patterns and create models which attempt to capture effective behaviour. In this case, the implications of having accurate models extend far beyond the fields of psychology and economics, as the expectations one has about individuals' actions, and reactions to different stimuli, closely influence the way we deal with them. The models we adopt have a drastic impact on public policy, since they are designed with an idea about human behaviour and compliance in mind (Sibony and Helleringer 2015).

For a long period of time, the rational choice model was used to make predictions about human behaviour and to inform policy decisions. It has, however, been widely criticised for making "systematic errors in predicting behavior" (Thaler 1980). Kahneman and Tversky's (1979) research challenges both these predictions and the policies which were created with rationality as its basis, by showing that these predictions are not true in specific conditions. They also offer an alternative since the diversions from these predictions are themselves also predictable, and paint a more accurate picture of human behaviour. In this sense, the concept of irrationality does not have the meaning it is usually given (namely, that a certain behaviour is unpredictable), but simply means that it is not expected by the rational choice model, while still being predictable. These new insights offer more accurate predictions which lead to an improvement of the design of public policies (Ariely 2008; Benartzi et al. 2017; Chetty 2015; Congdon, Kling and Mullainathan 2011; Loewenstein

et al. 2012; Sibony and Helleringer 2015; Thaler and Sunstein 2008).

Reviewing the impact of behavioural economics on public policy, Chetty concludes that "behavioral economics can offer more accurate and robust prescriptions for optimal policy" (2015, 37). In the same way that Friedman was arguing for positive economics, focusing on the impact of models' impact and predictive power rather than their assumptions (Friedman 1966), Chetty argues that the inclusion of behavioural sciences within economics often yields better results than traditional policies, and can also improve the evaluation of existing policies. These improvements have been explored and tested throughout the last decades (Benartzi et al. 2017; Bowles 2008; Bowles and Hwang 2008; Goldstein, Cialdini and Griskevicius 2008; Johnson and Goldstein 2003; Johnson et al. 2013; Madrian and Shea 2001; Mathios 2000; Ohman, Flykt and Esteves 2001; Schwartz et al. 2012; Selley and Sanders 2017; Shampanier, Mazar and Ariely 2007; Skov et al. 2013), and led to the development of new policy tools such as nudges. The next section focuses on the literature on nudges, as well as their use, and support for the policy.

1.2 Nudges, empirical use, and support

Nudges use behaviourally informed predictions in order to modify individuals' choice architecture (CA) and lead them towards one choice rather than another (Thaler and Sunstein 2008). As new policy tools which are thought to be more effective than traditional policies (Benartzi et al. 2017; Johnson and Goldstein 2003; Thaler and Sunstein 2008), nudges have quickly gained prominence both in academia, and with governments (Economist 2017).[3] The use of nudges raises many questions ranging from their efficiency to ethical considerations which arise when cognitive biases are used to stir individuals' decisions, and has

[3]Cass Sunstein has famously advised former United States of America President Barack Obama as administrator of the Office of Information and Regulatory Affairs, and Richard Thaler helped the United Kingdom's Cabinet Office to set up the Behavioural Insights Team which has been nicknamed the "nudge unit."

sparked a lot of interest from academics. In this part, I review some of the important literature on nudges and ethical considerations, before giving several examples, and presenting empirical research assessing whether people welcome the use of nudges as public policies.

1.2.1 Nudges: Definition and justification

Thaler and Sunstein coined and popularised the term 'nudge' in their seminal *Nudge: Improving Decisions about Health, Wealth, and Happiness* (2008). In the book, they describe ways in which individuals make decisions that are not optimal according to their own goals, and propose the use of nudges in order to help them accomplish their goals adequately (Thaler and Sunstein 2008, 79–88). As the subtitle of the book implies, they explain nudges as helping individuals make better decisions for themselves, improve their "health, wealth and happiness" and counter cognitive biases which lead to decisions which are detrimental to them (Haushofer and Fehr 2014; Shah, Mullainathan and Shafir 2012). Although Thaler and Sunstein do not have a precise technical definition of nudges (Hausman and Welch 2010; Saghai 2013), Saghai summarises their operationalisation of the concept as "A nudges B when A makes it more likely that B will φ, by triggering B's automatic cognitive processes, while preserving B's freedom of choice" (2013, 487). In other words, B is being nudged when their CA is modified in a way which (i) preserves their freedom of choice (e.g. by preserving all the options, and not using coercion) and (ii) uses their automatic cognitive processes to favour one choice over the others.

It is important to note that they later define nudges in terms of effect rather than intention, defending CA modification as "unavoidable" as "[n]ature itself nudges" (Sunstein 2015b, 11), but for this research, I focus on human-made nudges and ignore other 'nudges,' such as the weather. Ever since the publication of their book, the definition of nudge has been widely discussed, and several characteristics have been emphasised depending on the focus of the research or discipline. In an article reviewing results from policies using nudge,

Thaler and Sunstein simply refer to them as "behaviorally informed policies" (Benartzi et al. 2017, 1042), which is broader than their initial definition.

Nudges can take many forms and shapes, but always rely on specific cognitive biases, heuristics or rules of thumb that people use when making decisions (Thaler and Sunstein 2008). There are two very common examples which help understand the concept: the "cafeteria" (Thaler and Sunstein 2008, 1–3), and "urinal" nudge (3–4). The cafeteria nudge consists in modifying food placement in a cafeteria (Bovens 2008; Noggle 2018; Thaler and Sunstein 2008). Knowing that consumers are more likely to select what comes first and is easiest to see, placing healthy foods and meals at the front of the display increases the number of healthy meals which are purchased and decreases unhealthy ones (Thaler and Sunstein 2008). Consumers indeed fill their trays with the first dishes they see, and have already selected food by the time they arrive in front of the less healthy meals. This example shows that individuals' freedom of choice is preserved in the sense that they can still select cakes instead of salads, and the costs of doing so—the costs of resisting the nudge—are so small that they can be ignored.[4]

Concerning the urinal nudge, Amsterdam Schiphol airport printed flies at the bottom of urinals in order to 'give an aim' to urinal users (Thaler and Sunstein 2008).[5] Thanks to this visual cue, users subconsciously aim at the fly while urinating—in other words, they aim inside the urinal. Using this nudge, the airport was reportedly able to reduce spillage by 80%, and dramatically reduce cleaning and maintenance costs. This nudge also preserves freedom of choice (although urinating outside the urinal is arguably not a choice but rather the result of absentmindedness), and shows more clearly the reliance on automatic cognitive processes present in Thaler and Sunstein's (2008) and Saghai's (2013)

[4]If one does not agree that the costs of reaching further to get cake is negligible, they can consider Noggle's example in which the cake plates are made slightly easier to pick up than the salad bowls in order to offset the costs (2018, 165).

[5]There are alternatives to the use of flies, such as other prints or the placement of a little ball inside the urinals, but the logic is the same.

definitions. The nudge indeed solely relies on the fact that urinal users will automatically look, and therefore aim at, the fly inside the urinal.

The concept of nudges, as well as the legitimacy of its use has been widely discussed and challenged. Thaler and Sunstein inscribed the use of nudges within the idea of 'libertarian paternalism' (LP), famously claiming that it is not an oxymoron (Sunstein and Thaler 2003). The idea of LP is that regulators are libertarian paternalists when they "steer people's choices in directions that will improve the choosers' own welfare" (Sunstein and Thaler 2003, 1162). The paternalistic aspect of LP, which they call "relatively weak and nonintrusive" comes from the attempt to influence people's choices (Sunstein and Thaler 2003, 1162), while the libertarian aspect comes from the conservation of their freedom of choice. This approach has, however, been criticised for either not respecting liberal principles, or for being too paternalistic to be able to claim being libertarian.

Directly responding to Thaler and Sunstein, Mitchell (2005) claims that LP is indeed an oxymoron as it still means that the architect uses the nudgees' "irrational tendencies"—i.e. cognitive biases—in order to direct them in a certain direction (1276). While Mitchell (2005) agrees that since nudges do not modify the choice-sets, nudgees always have the possibility to avoid the nudge, he points out that the costs of resisting will always be borne by rational individuals, or individuals who are making an effort to be rational (Rizzo, Glen and Whitman 2009). However, as Dworkin (2013) explains, there is a difference between a possibility, or the capacity, to resist, and the exercise of that capacity. This means that the costs of resisting, even if arguably very low, might effectively prevent individuals who would want to resist the nudge, to do so. Regardless of the supposed resistibility of nudges, many scholars identify an issue with the fact that nudges still involve an individual, or agency, deciding what is important for the nudgee, and what will improve their welfare (Guala and Mittone 2015; Hausman and Welch 2010; Lodge and Wegrich 2016).

Grüne-Yanoff (2012), Guala and Mittone (2015) and Rizzo, Glen and Whitman (2009)

focus on the impossibility for architects to accurately make choices for the nudgees and their welfare. One of the justification for nudges is indeed that it improves the nudgees' welfare, but Grüne-Yanoff (2012) claims that it does not qualify as increasing welfare since the approach ignores the nudgees' values when it nudges them towards a choice. Guala and Mittone (2015) conclude that nudges cannot be justified from a welfaristic standpoint, but justify the use of certain nudges because they lead to a better society for us and future generations. Grüne-Yanoff (2012) concludes that LP violates liberal principles because the choice architects restrict nudgees' freedom by making one choice more likely to be chosen.

Lodge and Wegrich's (2016) and Rizzo, Glen and Whitman's (2009) critiques, however, are aimed at the rationality of the nudge approach and the architects. Policymakers and choice architects in general are prone to their own cognitive biases which might negatively impact the results of a nudge programme (Lodge and Wegrich 2016). The overconfidence bias, for example, which results in individuals over-estimating their own judgement relative to their actual accuracy, is very likely to affect a choice architect.

Proponents of nudges furthermore implicitly claim that they are able to identify many characteristics about the nudgees, which Rizzo, Glen and Whitman (2009) argue is impossible. Amongst others, these characteristics are: individuals' true preferences, the extent of the presence of a specific bias, self-debiasing methods already used by individuals, the possible interdependence of biases, as well as how nudges might influence nudgees' future self-regulation mechanisms (Rizzo, Glen and Whitman 2009, 966).[6] They conclude that, considering the lack of information about individuals' preferences and reactions to nudges, there is a need to put the use of nudges on hold (Rizzo, Glen and Whitman 2009).

The literature on nudges also focuses on their transparency and publicity. Hausman and Welch (2010) argue in favour of publicity in order to inform citizens about the means employed to change their behaviour and influence public policy. Although they do not

[6]Most of these concepts are self-explanatory, but Rizzo, Glen and Whitman explain them with great details.

advise against the use of nudges, they argue that since it is important for individuals to understand which of their choices are being influenced and why, it is important that governments couple their use of nudges with rational persuasion. Thaler and Sunstein (2008) rely on Rawls' publicity principle in order to prevent citizens from being manipulated by the government through nudges, because they assume that disclosure would prevent such manipulation. Hansen and Jespersen (2013) and Hansen, Skov and Skov (2016) focus on the transparency of nudges, and create a framework to assess their usability depending on their levels of transparency. For this framework, they distinguish between nudges that rely on reflective or automatic behaviour, and whether they are transparent or not.

1.2.2 Some empirical examples

Since the concept has appealed to so many researchers and policy-makers, nudges have been applied and tested in different fields. In order to have some examples in mind about the type of nudges which are empirically used, this section presents several studies which use and test the effects and efficiency of nudges in the areas of social norms and health.

In a field experiment which attempts to nudge hotel clients into using their bath towels more than once, Goldstein, Cialdini and Griskevicius (2008) test the efficacy of several messages. They found that the use of messages referring to social norms, for example emphasising that other guests have reused their towels, yielded better results than messages focusing on environmental protection. They also found that the closest social norms had the highest effect, so telling people that previous guests in their room reused their towels had a greater effect on their behaviour than telling them that other guests in the hotel did. Reviewing several studies which focus on the use of social norms, John, Sanders and Wang (2014) found that most studies recorded an effect in the expected direction. They, however, note that the results were modest (between 2 and 5%) and different depending on the fields, energy consumption showing the largest effect. Nagatsu (2015) explains the

performance of social nudges by explaining that they create an expectation for cooperation and bring individuals away from the "I frame," and closer to the "we frame" (489).

Using nudges for health issues is very relevant in the sense that it can provide great benefits to the nudgees without much costs. Moreover, individuals' best interest in terms of health seem, at least at first sight, to be easier to identify than in other areas, as the assumption that people generally want to be healthy would not be contested by many.

In a study which became a major example of the use of default options, Johnson and Goldstein (2003) investigated the possibility to get more people to sign up as organ donors. Noticing that while 85% of people in the United States approve of organ donation, only 28% were registered as donors, they investigated whether using the option to be a donor as the default option would increase the number of donors. As the default option is the one individuals will sign up to if they do not change the initial setting, people's tendency for inertia leads them to leave the default option selected, especially when they agree with the issue. Johnson and Goldstein (2003) find that using 'being a donor' as the default option would increase the number of both donors and donations. Investigating whether default options still work if people are informed that they are being nudged into making one choice over another, Loewenstein et al. (2015) find that there is no substantially or statistically significant difference in effect size between studies which inform and do not inform their participants.

Nudging for health mostly focuses on small changes which, repeated, can have a drastic impact on people's health (Blumenthal-Barby and Burroughs 2012; Hallsworth 2017; Hansen, Skov and Skov 2016). Many of the studies focus on food consumption, and alter participant's CAs in a way similar to the cafeteria nudge described by Thaler and Sunstein. As explained in the introduction, Mathios (2000) observed the effect of fat disclosure on salad dressing packages, and concluded that people change their behaviour towards lower-fat products when they are provided with the information. Also focusing on the

design of labels on food products, Roberto and Kawachi (2015) and Roberto and Khandpur (2014) give recommendations to improve the design of food labels in order to better inform consumers. In an experiment in a fast food restaurant, Schwartz et al. (2012) were able to reduce calorie consumption by more than 200 units per meal by asking customers if they wanted to downsize their side dishes.

Such results are exciting as we know that a consistent calorie consumption reduction of even 200 calories can have great impacts on individuals' lives (Hallsworth 2017). In other experiments, however, the reduction in calories resulting from the nudge were either offset by individuals ordering more calories in another area (Wisdom, Downs and Loewenstein 2010), was not statistically significant or did not lead to the expected changes in behaviour (Loewenstein 2011; Loewenstein et al. 2012). Loewenstein et al. (2012) conclude that behavioural policies cannot be used as substitutes for traditional policies in order to promote health, and that they ought to be used to compliment them. Burgess (2012) adds that nudges do not contribute nor encourage individuals to make better decisions, but only punctually influence them to do so. This means that they do not encourage a better decision-making process, but only push people to make certain decisions. Furthermore, nudging individuals for their health comes with a bundle of ethical issues which are not often discussed nor considered (Blumenthal-Barby and Burroughs 2012).

This list of studies reviewing the use of nudges is in no way exhaustive (even for the two topics I have selected), but such a list would be too long to compile, and to read. Some reviews are however more extensive than mine, such as Sunstein's (2011) which both explains and gives examples of behaviourally informed policies. Two other papers which do not only present nudges, but evaluate their results against other policies are provided by Benartzi et al. (2017) and Chetty (2015). A short review is provided by Sunstein (2014), as well as a list of what he calls "ten most important nudges" (abstract). Following this presentation of nudges, I review empirical research on whether they are accepted by citizens.

1.2.3 Do people like nudges?

Sunstein, who has prolifically written to promote nudges, conducted several opinion surveys across the world in order to ask people whether they approved of the use of nudges by their governments (Reisch and Sunstein 2016; Sunstein 2015a, 2017a, 2017b). In democratic societies, while the content of public policies is (and should) always be debated, the way the government carries these policies out is equally relevant although less debated in the public eye. It is therefore important to know whether individuals approve of their governments' use of nudges as policy tools.

Starting with North America (Sunstein 2015a), and expanding their research to Europe (Reisch and Sunstein 2016), and the rest of the world (Sunstein 2017b), Sunstein found that most people like nudges which promote legitimate goals in ways they see legitimate. With different levels of support across countries and regions, the surveys find the same patterns in terms of the nudges people approve and disapprove of, and it seems that people tend to like nudges in general, but dislike the ones which are seen as overly manipulative, such as subliminal images for example.[7] Furthermore, it seems that nudges do not only lose support when they are seen as overly manipulative, but also when the method employed is non-manipulative but the outcomes are not consistent with respondents' interests and values (Sunstein 2017a, 2017b).

Sunstein, Reisch and Rauber (2017) categorise countries in three groups depending on their overall level of acceptance for nudges. These three groups are (i) countries where strong majorities approve of nudges which fit their interests and values (mostly liberal democracies), (ii) countries with overwhelming majorities in favour of nudging (China and South-Korea), and (iii) countries with markedly lower approval ratings (Hungary, Denmark and Japan) (Reisch and Sunstein 2016). Sunstein (2015a) points that an important component of such

[7] This finding fits the discussion Thaler and Sunstein (2008) have about this method.

acceptance rates is the transparency of the government's decision to nudge and the types of nudges themselves.

This section has introduced the concept of nudges, as well as the literature which supports it and empirical research which looks at nudges' acceptance by the general public. The next section reviews a part of the literature on advertising which focuses on advertisement and its manipulative aspects.

1.3 Advertisements and choice architecture

As this thesis is concerned with assessing whether nudges can be used in order to counter the manipulative aspect of advertisement, it is important to understand advertisement, and its reliance on choice architecture (CA) modifications. As explained in the previous sections, nudges, and the use of CA modification for public policy has been popularised fairly recently. Advertisement, however, has relied on modifying customers' CAs since its inception. The goal of advertisement, whether it is for a brand or a product, is to change the way customers perceive their alternatives (Packard 1957). Advertisers' goal is to change, or create, the vision customers have of certain products in order to increase sales. When we are in front of unknown brands in the biscuit section of a supermarket in a country we do not speak the language of, our CA is entirely created by what the brands (advertisers) want us to know of their products. In this case, our CA is entirely constructed, and we are victim to the way the products are presented and packaged. We are manipulated into buying the biscuits whose packaging makes us think it is the best product.

This situation does not happen every day, and what happens most of the time is far from a CA being entirely constructed by advertisements.

Such a situation, in which a CA is entirely constructed by advertisements, is, however, very uncommon. Instead, they are simply modified by messages, manipulating us to buy

certain products (Beauchamp, Hare and Biederman 1984; Danciu 2014; Main 2012; Phillips 1997; Shabbir and Thwaites 2007). Although these have not been formally studied and empirically tested, advertisements have long relied on the same cognitive biases which now give policy makers the ability to help individuals make better choices. Advertisement indeed also makes use of framing (Smith 1996; Treise et al. 1994), and priming effects (Wänke 2016), as well as the use of social rules and deception (Danciu 2014).

Cognitive biases and CA modifications are used with the intention to change customers' vision of products or brands, and often attempts to associate them with a specific desire (Phillips 1997; Smith 1996). A great example is the Marlboro tobacco company advertising their cigarettes by framing their product as enhancing freedom and representing the spirit of adventure. There is a diversion away from a product's pure characteristics, to what it represents (Zhang et al. 2014), which means that customers do not buy cigarettes but buy freedom and adventure instead. The use of this type of associative advertisement is becoming more widespread as brands increasingly attempt to associate their products or themselves with a lifestyle, or an ideal. Other examples are Apple advertisements associating their products with a way of life, or Coca-Cola's "Open Happiness" and "Taste the Feeling" campaigns.

Focusing on neuro-marketing and observing how our brain reacts to advertisements, Main (2012) argues that many of the advertisement techniques which are used create an unfair advantage on the side of the companies because they exploit biases and are hard to resist. Because of these techniques, and despite the fact that some advertisement techniques might not be, advertisement is very often referred to as manipulative (Beauchamp, Hare and Biederman 1984; Shabbir and Thwaites 2007; Treise et al. 1994). Phillips (1997) opposes manipulative to informative advertisements, and argues that manipulative advertisement constitute around half of all existing advertisements. Whether this number is correct is another question, but the use of informative advertisement has certainly decreased since

1997, when he wrote.

When it comes to CA modifications through advertisement, it is important to note that there are several layers to consider. In addition to being modified by the way products are framed, made to represent, and packaged, our CAs are also modified by the way products are placed when we see them. Following the same mechanism as the cafeteria nudge, the placement of products in supermarket also influences customers' choices (Escaron et al. 2013). For example, the cheapest products are most of the time situated at the very bottom and are harder to see. Similarly, having a very expensive appeal product (a television for example) next to other ones also modifies the way customers see their alternatives, and leads them to comparatively perceive the alternatives as cheaper (Shampanier, Mazar and Ariely 2007).

After explaining how advertisement manipulates consumers, Danciu (2014) gives the recommendation to create "conscious consumers" in order to counter this manipulative aspect (29). Choices which are induced by manipulative advertisement are indeed rarely beneficial to consumers as such, as private companies advertise to boost their revenues, and not their customers' welfare. There are exceptions in which a company's sale goals are in line with their consumers' welfare, but they are seldom enough to ignore here. As Danciu's recommendations focus on changes within the advertisement industry, he does not address the use of nudges which is explored in this thesis.

1.4 Manipulation, autonomy, and behavioural sciences

The last stream of literature which is relevant to the analysis done in this thesis relates to manipulation and autonomy. Both of these concepts are extremely important when discussing both nudges and advertisement. Proponents of nudges can claim that their use is libertarian paternalist as it preserve freedom of choice and thus do not violate

individuals' autonomy. This does not necessarily hold true when it comes to advertisement, as its manipulative aspect is by definition a violation of individuals' autonomy in the sense that it moves their desires away from their ideal settings (Noggle 1996). In this section, I review a small portion of the literature on manipulation which has touched upon nudges and more generally on mechanisms relying on cognitive biases.

Noggle (1996) defines manipulation as "the attempt to get someone's beliefs, desires, or emotions to violate" their "norms or ideals that govern beliefs, desires, and emotion" (44). He argues that this is done through the adjustment of individuals' "psychological levers" with the intent to push them to make a different decision than they otherwise would have (Noggle 1996, 44). The three levers he defines—belief, desire, and emotion—can be used in order to manipulate people in different ways through different mechanisms. In this sense, to manipulate someone is to intentionally use their psychological levers in order to move their judgement away from its usual settings.

By inducing someone to make a decision they would not have initially made, thereby effectively manipulating them, the manipulator interferes with their independence and autonomy, and "undermines the agency of the manipulated party" (Moles 2015, 656). As a source of reduction of individuals' autonomy, manipulation is therefore not enviable, but we can hardly argue that manipulation should be prevented altogether. Moles (2015) indeed notes that there is a "presumption against manipulation [...] rather than a prohibition" (656–7). He provides an example in which a child manipulates her father by exaggerating her crying in order to receive another piece of cake (Moles 2015, 655). Few would argue that this type of manipulation should be prevented, and this is because the action the father was manipulated into doing was not against his interest.

Although it is *prima facie* objectionable, individuals would not object to all cases of manipulation. In fact, several nudges included in Sunstein's research about acceptance qualify to this definition of manipulation but were widely accepted because their ends fit

the respondents' goals and values. Advertisements, on the other side, are not in line with consumer's interests, and qualify as the type of manipulation which is objectionable and hinders individuals' autonomy in a way which should be countered.

Scholars interested in manipulation have naturally also been interested in nudges, and vice-versa (Moles 2015; Noggle 2018; Nys and Engelen 2017; Wilkinson 2013), but very few have focused on using nudges as counter-manipulation mechanisms. Wilkinson (2017), for example, observes the case of counter-manipulation for health promotion with some arguments close to the ones for nudges, but does not explore the idea. He is sceptical about counter-manipulation mechanisms in general because he argues that they result in a loss of autonomy. He, however, writes that solving the autonomy objection would make counter-manipulation more attractive as it would no longer imply reducing individuals' autonomy (Wilkinson 2017, 265). He only considered traditional counter-manipulation policies, so the use of nudges might be a good solution to solve the autonomy objection.

Glod (2015) argues that many nudges are often problematic because they do not respect people's autonomy and fail to identify their preferences. Although he is not explicit about it, his position does imply that they would not be problematic if they did respect individuals' autonomy. Wilkinson (2013) writes that since manipulation requires the intention to manipulate, most nudges are not manipulative, and can even be consistent with individuals' autonomy if proper consent is given. This would mean that in the case of these nudges, even Glod would agree that they do not violate autonomy. Focusing on health nudges, Nys and Engelen (2017) further argue that many nudges can be justified as they help individuals increase or maintain their autonomy. Furthermore, even if they violate individuals' autonomy, some nudges might be justified and accepted by a democratic procedure.

Considering nudges which change the salience of some choices over others, Noggle (2018) distinguishes two types of manipulation, namely as pressure or as trickery. He concludes that nudges cannot be qualified as pressure since they do not exert pressure on nudgees,

however, the ones who rely on disproportionately altering the salience of some choice can be qualified as trickery. This means that in order to consider nudges as counter-manipulation mechanisms, one needs to be wary of their impact on individual's autonomy, the mechanisms they use, as well as their justification.

An important aspect of manipulation relevant to the discussion, is that manipulation "has a success condition" (Wilkinson 2017, 258). This means that for an action to be manipulative, it is not sufficient that an agent acts manipulatively towards another, but the target's judgement must have been successfully swayed away from its usual settings. This has implications for both nudges and advertisement, as it means that neither can be considered manipulative unless they succeed in changing individuals' behaviour, and this even if they act manipulatively—that is with the intention to manipulate (Wilkinson 2013).

Reviewing the literature on advancements in behavioural sciences helps in understanding the concept of irrationality, and how diversions form the rational choice model can be used to modify people's behaviour. Relying on these diversions, both nudges and advertisements modify people' CA to trigger a specific behaviour and achieve their respective roles. On the one hand, nudges attempt to increase individuals' welfare, or in certain cases, to prevent them from making decisions which would reduce it. On the other hand, advertisements modify people's CAs in order to sell more of the products they are advertising.

Chapter 2: Argument

To determine whether nudges can be legitimately used as a counter-manipulation mechanism, I need to answer several questions raised by the literature. I argue that nudges can generally be used as counter-manipulation mechanisms against advertisement, but some limitations need to be observed concerning the content, transparency and publicity of the nudges used.

I first present the boundaries of the argument, outlining what is considered for nudges and advertisement. In a second section, I focus on the justification needed for using nudges to counter manipulation by advertisement. The third section distinguishes nudges based on whether they are in line with nudgees' preferences and their transparency and explores whether governments can use nudges to counter manipulation.

2.1 Boundaries of the argument

Before diving into whether, and how, nudges can be used to counter manipulation by advertisement, I need to establish the premises of the discussion. This section offers a characterisation of nudges, and explains that advertisements are taken as manipulative. Building on the relevant literature, I take nudges to be intentional, relying on cognitive biases and easily resistible.

2.1.1 Nudges and characteristics

As explained in the previous chapter, nudges lack a clear, universally accepted definition (Hausman and Welch 2010; Saghai 2013), which means that several studies use the concept slightly differently. This conceptual unclarity has not led to any misunderstanding, but often leads to policy interventions being described as nudges by some, while they would

not fit other definitions previously defended. Some of the nudges described by Benartzi et al. (2017), for example, such as "eliminat[ing] or redu[ing] paperwork requirements for obtaining licenses" (1042), would not fit a definition of nudge which focuses on "automatic behaviour" such as the one presented by Saghai (2013, 487) and Thaler and Sunstein (2008).

It is therefore important that I stipulate characteristics which define nudges. For the purpose of this research, I take nudges to have three main characteristics. Nudges are (i) human-designed, and intentional, choice architecture modifications which (ii) rely on individuals' cognitive biases to make them take one decision over another, and are (iii) easily resistible.

The first characteristic comes from the fact that Thaler and Sunstein define nudges as any CA modification, whether it is human-made or not, and whether it is intentional or not (Sunstein 2015b; Thaler and Sunstein 2008). While they are right to explain that any CA modification, whether intentional or not, can modify behaviour, they are not all relevant to evaluate in terms of their ethics. While the fact that the sunny weather in Nice today modified my options and led me to only wear a shirt and leave my coat at home would qualify as a nudge as defined by Thaler and Sunstein, it is not relevant to observe when considering whether governments can use nudges as policy tools. The nudges I consider therefore need to be intentional. Since there is no neutral CA, without this characteristic, any CA creation and modification could potentially qualify as a nudge.

The second characteristic is what differentiates nudges from other types of policies or interventions. While other policies or interventions rely on traditional economic mechanisms such as financial incentives (Benartzi et al. 2017; Bowles and Hwang 2008; Chetty 2015), nudges use individuals' diversions from the rational choice model to change behaviour. As for the third characteristic, the resistibility, is needed in order to make sure that fully rational individuals are not affected by the nudge. Nudges are used in order to correct for

individuals' deviations from rationality—helping them to make the choice they would make if they were fully rational—and not to change rationally chosen decisions.

2.1.2 Advertisement

Defining advertisement is more straightforward than defining nudges, but not all forms apply to this research. An advertisement is generally thought of as a message (which can be on any medium) or action which has the goal to promote a product, service, brand, event, person, belief or ideology. For this thesis, I make two further distinctions between messages and actions, and between manipulative and informative advertisement, and only focus on messages and manipulative advertisements.

The first distinction is needed because advertisement in the form of actions cannot realistically be nudged against. Let us imagine a situation in which a traveller walking out of a train station is handed over a can of fizzy drink by a worker dressed in the colours of the drink's brand. At that moment, any action which would prevent the voyager to take the can would have to be coercive and physically prevent the exchange, which means that nudges are ruled out. The only influences of this action-advertisement which could be countered using a nudge are its longer term effects induced by brand recognition. These effects will be dealt with in the following section.

The second distinction, between manipulative and informative advertisement, is one which has been made by Phillips (1997). Informative advertisement, on the one hand, relates to advertisement which mainly focuses on informing consumers about the characteristics of what is advertised. Manipulative advertisements, on the other hand, do not use the product's or service's characteristics to promote them. While it is unlikely that an advertisement is fully informative, only showing characteristics, or fully manipulative, not even showing the product, they all focus on either side. Manipulative advertisements therefore are the ones which mostly rely on non-characteristics elements to promote what

they advertise. Examples of this type of advertisement are not scarce, and good examples are perfume advertisements. It is, however, important to keep in mind that Phillips (1997) uses the term 'manipulative' to refer to the use of manipulative acts, and not to the success of these manipulative acts. I only consider manipulative advertisements which are actually successful in manipulating their target.

From now on, I use the word advertisement to mean a (i) promotional message used to increase sales for a product, service, brand, event, person, belief or ideology which (ii) does not focus on the characteristics of the product, service, brand, event, person, belief or ideology, and (iii) is successful in manipulating its target. Furthermore, many examples in this thesis relate to television or internet advertisements, but it is important to note that the argument, as well as the use of nudges to counter manipulation, applies to advertisement distributed through every media.

2.2 Justification for nudges

Now that I have set the boundaries of the argument, I move on to explore the justification for such a use of nudges. I first consider whether nudges differ from traditional policies, and whether such a difference could lead to nudges needing more justification than traditional policies. If nudges do not require more justification, it would mean that they can already be used by governments in areas in which traditional policies are used. After concluding that no further justification is needed, I present two ways in which nudges can be used to counter manipulative advertisement.

2.2.1 Do nudges differ from traditional policies?

Traditional policies, such as incentives, educational programmes, and regulations are usually justified by governments' decision making procedures. In democracies, policies which are

implemented by governments and/or parliaments are justified because they are indirectly chosen by the electorate. For nudges, however, the fact that they direct people towards a choice and that they rely on individuals' cognitive biases seems to call for, at least *prima facie*, further justification. I will explore both claims, before concluding that most nudges do not need more justification than traditional policies.

Reviewing the literature on behavioural sciences leads to the conclusion that traditional policies are generally wrong in their expectations of human behaviour. They do not have the desired effect because they expect citizens to act rationally (Alemanno and Sibony 2015; Alemanno and Spina 2014; Sibony and Helleringer 2015). Because of this assumption, policy makers have thought that many details of the design of policies did not influence citizens' compliance, and we now know that this is not the case (Thaler, Sunstein and Balz 2012). Thaler, Sunstein and Balz (2012) indeed write that "[i]n a world of Econs [rational individuals], these details [the way choices are organised] would not matter, but for Humans, nearly everything matters" (430).

Acknowledging that 'nearly everything matters' when it comes to individuals making choices, means that traditional policies always modify CAs, thus directing individuals' choices. As Thaler and Sunstein (2008) point out, there is no "pure neutrality" and any CA modification will have an impact (246). The remaining difference between traditional policies and nudges, when it comes to directing individuals towards a choice, seem to be the intentionality of the CA modification. The former direct individuals towards a choice without necessarily realising it, while the latter are designed to favour some choices over others.

Following from this intentionality, a common argument against nudges is that they are paternalistic in the sense that they direct individuals towards choices that the architects deem better for the nudgees than what they would otherwise choose. In this case, the argument might be used in favour of traditional policies, arguing that even though they

direct individuals towards a choice, they do not do it in a paternalistic way since they do it unintentionally. This does not, however, apply to all since some traditional policies are clearly paternalistic and criticised as such, as is the case for sin taxes targeting tobacco or alcohol. There are three things to consider about this objection, and paternalism in general. First, some traditional policies are also paternalistic, second, a nudge is not paternalistic if it directs people towards choices in line with their preferences, and third, nudges are easily resistible (often even more so than traditional policies).

Many traditional policies are also paternalistic in the way they modify CAs. Policies which ban advertisement for unhealthy food on television during children programmes, or for cigarettes and other health damaging products, which exist in many countries, are indeed paternalistic. However, they are not paternalistic when they direct people to make choices in line with their preferences. Such a ban will not be paternalistic towards an individual who has a preference towards not smoking, but will be towards someone who wishes to smoke.

This is also the case for nudges. They are only paternalistic if the architects direct individuals towards a choice which is not in line with their preferences (I discuss preference alignment in the next chapter). In other words, one which they would not have taken if they were rational decision-makers. They can, however, still be paternalistic. If someone with a preference for chocolate cake is nudged into choosing a salad in the cafeteria because of the owner's arrangement choices, the nudge would be paternalistic. However, since I define nudges as easily resistible, even if they have the potential to be manipulative, they will likely not be. People with a preference for chocolate cake can simply walk further without taking a salad and pick up a slice. Nudges are not necessarily paternalistic, and even if they might be in some situations, the fact that they are resistible undermines the possibility.

Furthermore, it is arguable that some traditional policies are even harder to resist, if

they are resistible at all. The use of financial incentives, for example, might be difficult to resist, especially if the reward is high. This is principally the case for individuals from lower socio-economic backgrounds, which means that the resistibility of these policy would only be available for specific parts of society, but this aspect is beyond the scope of this thesis.

The fact that nudges direct individuals towards a specific choice does not constitute a diversion from traditional policies which would require further justification. Indeed, both types of policies modify CAs, and the fact that these modifications are either intentional or not does not lead to a difference in the degree of paternalism which would require justification. I will now question whether the fact that nudges rely on cognitive biases requires specific justifications.

Their reliance on people's cognitive biases seems to bring a *prima facie* presumption against nudges which would lead to require extra justification. This presumption comes from the fact that 'relying on cognitive biases' makes nudges wrongly seem as if they were changing the way our brain works.[8] This is however, not entirely the case. As I have explained earlier, cognitive biases are only diversions from the rational choice model. In other words, they are a diversion from what was arbitrarily used as an ideal version of human behaviour and decision making.

Cognitive biases are not biases in the sense that they are mistakes from an actual ideal decision-making process, but from a model of this decision-making process. As was pointed out by research in the behavioural sciences, this model is erroneous and does not capture the way we actually make decisions. Cognitive biases are therefore only called biases, and do not point out errors in behaviour, but errors in the rational choice model's predictions of this behaviour. This means that nudges' reliance on cognitive biases simply signifies a switch to another understanding of human behaviour.

[8]If I did not know what cognitive biases are and someone would tell me that they will use them to change my behaviour, I would surely envision a Clockwork Orange scene.

Nudges and traditional policies simply use a different model of human behaviour in order to plan policy interventions. Both traditional policies and nudges need justification to be implemented. Traditional policies, however, do not require specific justification for their confidence in the rational choice model, and similarly, nudges should not need to justify their use of a different one. Nudges' reliance on cognitive biases alone does not call for further justification. It is, however, clear that there is a difference in the way they are perceived, which means that although further justification may not be required, architects need to be wary about their acceptance, and should therefore be strict about transparency.

After scrutiny, neither of the two characteristics—directing people towards a choice and relying on cognitive biases—which seemed to require more justification actually do. While I have discussed nudges generally, there seem to be differences between different nudges and contrasting situations in which they are used, depending on whether they are in line with the nudgees' preferences, and their transparency. I now explore how and whether governments can use nudges specifically to counter manipulation from advertisement.

2.2.2 Nudges versus advertisement

I identify two ways in which a government can use nudges in order to counter manipulation from advertisement. They can either make the manipulation less effective by changing the salience of the nudgees' options, or make it less relevant by arranging their CAs in a way which fits their goals and values regardless of advertisement.

The first way in which nudges can be used is by adjusting the salience of various options according to objective criteria, instead of images used in advertisement. These nudges reconcile the "salience and importance" pertaining to the products or services it targets (Noggle 2018, 166). Such nudges would, for example, modify the frame of medical choices focusing on survival instead of fatality rates (Bullock 2014), highlight the fat or sugar

content of products (Mathios 2000), or more generally provide nutritional information.[9] These nudges are simply attempting to make some information more salient than the ones advertisements initially highlighted. Returning to the example given earlier about Marlboro's framing of cigarettes as freedom, a counter-manipulative nudge would attempt to change consumers' focus from the 'freedom enhancement' idea to the health risks associated with smoking.

The second way is to use nudges regardless of advertisement, in order to make sure that nudgees' CAs fit their goals and values. The cafeteria nudge is a good example in which, regardless of which manipulative advertisements a customer has been exposed to, their CA is organised in a way which makes them more likely to choose an option in line with their goal to be healthy. Another example would be the arrangement of food in supermarket aisles. Supermarkets putting healthy foods in more prominent spots than processed foods (Escaron et al. 2013), and having recipe boards or leaflets next to fresh products would promote home cooking and healthier consumption (Papies et al. 2013).

Both ways can counter manipulative advertisement, whether directly as the first one, or indirectly, through the use of more general CAs, as the second one. Now that I have established a base to generally justify the use of nudges, and explained how they can be used to counter advertisement, I emphasise several characteristics of nudges and assess whether, and how, they can be used by governments.

2.3 Characteristics of nudges

Until now, I have mostly considered nudges as a broad category instead of differentiating individual nudges. The concept of nudge, however, can include several different interventions which might need further explanation or justification. In this section, I explain how nudges

[9]For a more extensive discussion of salience nudges, see Noggle (2018).

can be different depending on whether they are in line with nudgees' preferences, and their transparency. These two dimensions are important because they are the ones which will determine whether governments can use nudges to counter manipulation by advertisement. I conclude the section by briefly discussing the implications of these two dimensions on such use of nudges.

2.3.1 Preference alignment

I have discussed preference alignment earlier when claiming that the resistibility of nudges prevents them from acting against nudgees' preferences. Here, I explain exactly what this means and the implications for the use of nudges. I conclude that although they are easily resistible and misalignment will not result in manipulation, the possibility of misalignment between nudgees' preferences and nudges calls for a strong focus on their transparency.

Nudges can therefore be either aligned or mis-aligned with nudgees' preferences. They are considered aligned when the choice they favour fits with the nudgees' preference, and are mis-aligned when they do not. The cafeteria nudge, for example, is aligned for consumers who have a preference for eating healthy, and mis-aligned for the ones who actually have a preference for eating unhealthy. It is important to note that the preference which is used to determine whether nudges are aligned or not is the one nudgees hold at the moment they are nudged. This means that for someone who has a general preference for healthy eating but wishes to indulge themselves with chocolate cake on the day they visit the cafeteria, the nudge will be mis-aligned. Whether nudges are aligned or not will therefore depend on the nudgee as well as the context.

A critique which is often brought forward against nudges and libertarian paternalism in general, is the impossibility for architects to know nudgees' true preferences (Hill 2007; Rizzo, Glen and Whitman 2009). Arguing against CA modifications because we do not know people's true preferences implies the thought that people would otherwise select the

option in line with their preferences, and that it is the modification which changes it away from their true preference. We know that this is not true and that people fail at identifying the option which is best according to their own preferences (Kahneman 2011). Since nudges are either aligned or not, there are two possible options and only one of them is an issue considering this objection.

Aligned nudges are never an issue since they do not act against people's preferences, regardless of what they are. This means that in the cafeteria, the fact that people with a preference to eat healthy take a salad is not a problem and simply means that the architects' assumptions about their customers were accurate. The reverse situation in which someone has a preference for eating unhealthy and is nudged into taking a salad, is however a problem which arises from architects not being able to perfectly infer people's preferences. This is a situation architects want to avoid, but it is not a situation which can arise from the nudges I explore here, since I have defined them as easily resistible.

With this definition of nudges, the two possible scenarios are therefore amended. Either the nudge is aligned, and the nudgee complies. Or the nudge is mis-aligned, and the nudgee resists the nudge, following their preference. In neither case, the fact that architects are not perfectly informed about the nudgees' preferences is an issue.

Much of nudges' benefits and robustness against objections therefore rely on the fact that they are easily resistible. In combination with the fact that they do not restrict the choice set—so leaving nudgees options to actually resist—resistibility is at the core of nudges' claim to libertarian paternalism. There is therefore a strong need to ensure that resistibility persists and is thoroughly built in every nudge which is implemented. In order for nudgees to be ensured that the implemented nudges are resistible, these need to satisfy some transparency requirements.

2.3.2 Transparency

Transparency is a very important characteristic of nudges, and this part reviews how nudges can be transparent. I define the possible transparency of nudges as either visible or transparent. A nudge is visible when nudgees can identify it but its aim is unclear and does not directly follow from its detection, and transparent when it is easily noticeable—or when it has been revealed—and its aim is easily identifiable.

Before describing visible and transparent nudges, it is important to note that nudge-like interventions which are invisible in the sense that they are not noticeable are not considered nudges. An example would be the use of subliminal images, which are not noticeable without the use of specialised equipment. Subliminal images are "invisible and thus impossible to monitor" (Thaler and Sunstein 2008, 264), and their content would still remain invisible in a way which would make it impossible for citizens to check that their content is actually about smoking reduction (Bovens 2008). Although they would technically act like nudges, this type of invisible intervention does not fulfil the resistibility feature of the definition and does therefore not qualify as nudges.

The first category is visible nudges. Visible nudges are noticeable in the sense that nudgees can reasonably notice that they are being nudged, but are not fully transparent as they cannot easily understand the nudge's goals. In this sense, the cafeteria nudge is to be considered visible. It is easy for a customer to realise that the healthy foods are placed at the beginning of the line while unhealthy foods are harder to reach. It is, however, not obvious that this was done for their health as it could be the result of a random process, the mood of the employee, or because these products are more expensive and bring more revenue to the cafeteria.

Since the goal of these nudges is not obvious at the moment when they influence the nudgees' behaviour, they require that their existence and goals be made public through a

different medium. This means that visible nudges cannot be implemented by governments if their use are not made public. I will not dive into specific requirements for publicity, but these nudges should be publicised in a reasonable way, in accordance to the national laws which apply to other policies. It is not necessary that governments run public campaigns to inform the population of their use of every specific visible nudge, but the information should be freely available to any citizen looking for it.[10]

The second category includes fully transparent nudges which nudgees can easily notice and identify the ends of. An example of a transparent nudge is one described by Thaler and Sunstein (2008) about lines drawn on the road ahead of a dangerous turn around Chicago (41). They describe that the lines get increasingly closer, giving the driver a sensation of increasing speed, making them slow down ahead of the turn. It is easy for drivers to see the lines, notice their effect, and understand why they have been drawn. Transparent nudges do not require extra publicity since the requirement is fulfilled at the moment the nudge modifies behaviour. This lack of necessity does not, however, imply that governments should not publicise them further, but simply that they do not need to.

One particular case to consider is the one of an intervention which would be publicised yet undetectable. Let us take the example of a government which publicises their use of 'nudges' as part of a campaign to reduce smoking. Citizens would be aware of both the fact that they are being nudged, and the reason why, but would remain unable to see it. In this case, as for the case of subliminal images discussed above, the intervention would not be qualified as a nudge. Even if they are known and that their goals are supported by the population, the fact that they are not detectable means that these interventions would essentially be "blank check[s]" to the government (Bovens 2008, 217). Publicity alone cannot justify nudges.

[10]In most cases, a publication on the government website, or the official journal would satisfy this requirement.

2.3.3 Can governments use nudges to counter manipulation?

To determine whether governments can use nudges to counter manipulation from advertisement, I need to further consider the content and transparency of the nudges used. Discussing content, I will consider their alignment with preferences, as well as the claim that nudges are manipulative.

There are two common objections to the use of nudges by governments. The first pertains to whether their content is in line with the nudgees' goals and values, and the second concerns whether they are manipulative in themselves. I have already discussed the content objection above when distinguishing between aligned and mis-aligned nudges, and it is not a concern government should have, since it is enough that they ensure nudges' resistibility and fulfil the transparency and publicity requirements. If these criteria are satisfied, mis-aligned nudges will be effectively resisted, and do not have to be considered a problem.

When it comes to manipulation, the claim is that nudges violate individuals' autonomy because they direct them towards a different choice than they would have otherwise chosen. This objection is similar to the one I entertained earlier about libertarian paternalism, concluding that, once again, their resistibility ensured that they did not result in individuals choosing options they would not have if they were fully rational.

In the case of counter-manipulation, Wilkinson (2017) considers that it can actually be respectful of individuals' autonomy. While he does not specifically consider nudges, the argument that counter-manipulation policies do not necessary violate individuals' autonomy applies to nudges and this discussion. Specifically about advertisement, I argue that counter-manipulative nudges actually increase nudgees' autonomy because they are meant to free them from the influence it would have had on their decision making.

Counter-manipulative nudges against advertisement safeguard nudgees' autonomy by

both making sure that they are not pushed to make a purchasing decision which would be against their interest, and by making sure they can take their own decisions in the long-run. In the short-term, counter-manipulative nudges prevent citizens to make decisions which are not in line with their preferences. This includes purchasing decisions which they would not make if they were not manipulated by advertisements. Such a nudge would, for example, have prevented me from being manipulated into buying a chocolate bar I saw in an advertisement at the film theatre yesterday, because it does not fit with my preferences of consuming less sugar.

In the long-term, counter-manipulative nudges would prevent people from being manipulated into making decisions which they would either regret or which would prevent them from making autonomous choices in the future. A straightforward example of such a decision is the case of financial decisions such as consumer credits, which can put individuals at strain in the long-run and prevent them from making their own decision later on. In this case, a counter-manipulative nudge would be effective at the moment of the initial decision, and prevent individuals from being manipulated into making a decision which will restrict their autonomy later.

Instead of reducing nudgees' autonomy, counter-manipulative nudges actually promote it by preventing that they are manipulated into decisions which do not match their preferences and can reduce their autonomy later on. Further considering salience nudges, which is the first way nudges can be used to counter-manipulation as explained earlier, Noggle (2018) argues that they will only be manipulative if they increase salience disproportionately relatively to "the thing's actual importance" (168). This means that the content of nudges is not a defining criteria for whether nudges can be used since more scrutiny is needed to determine whether they are manipulative or not. Additionally, even if some nudges might be considered manipulative, an objection against them is only *pro tanto* in the sense that a manipulative nudge might be accepted for its benefits regardless of its status

as manipulative. Because they might be accepted by the population for their benefits, the use of certain manipulative nudges by governments might be justified. Even if one considers nudges to be manipulative, this manipulative aspect cannot be used as a baseline for assessing whether nudges can be used, since whether a nudge's manipulative aspect is an issue for its usability depends on other factors.

In the previous part, I have emphasised the importance of transparency for both nudges' existence and goals. Ever since Thaler and Sunstein introduced nudges, transparency has been a key part of the literature on nudges, and for good reasons. Several studies argue for caution (Rizzo, Glen and Whitman 2009), and identify nudges depending on their level of transparency (Hansen and Jespersen 2013; Hansen, Skov and Skov 2016; Sunstein 2015b). I have made the distinction between visible and transparent nudges, and incorporated a need for publicity which is required for nudges' resistibility. If governments satisfy these requirements, they can use nudges to counter manipulation by advertisement.

I have argued that nudges can be legitimately used by governments to counter manipulative advertisement. After establishing that they did not require more justification than traditional policies, I have introduced two ways in which nudges can be used to counter manipulation from advertisement. I have also established criteria to define nudges, and set up transparency and publicity requirements for their use. In the next chapter, I entertain two objections to the use of nudges to counter manipulation from advertisement.

Chapter 3: Objections

There are two main objections to the use of nudges against manipulative advertisement, and I address them in this chapter. Firstly, critics might object to the need for counter-manipulation against advertisement. And secondly, if there is a need for counter-manipulation from advertisement, one might argue that banning advertisement altogether would be a better option than nudges. This chapter addresses both and concludes that nudges can, and should, still be used to counter manipulation from advertisement.

3.1 Why even try to counter manipulation?

The first objection does not directly challenge the use of nudges, but rather the need for counter manipulation in itself. Opponents of government intervention, for example, might argue that governments should not further modify citizens' CAs because it is paternalistic and intrusive. Although I have established that nudges do not require more justification than traditional policies, the objection is pertinent in the sense that individuals tend to be more wary of nudges than other more traditional policies such as incentives or educational measures.

The obvious reason for counter-manipulation in the case of advertisement, is to protect citizens against companies and prevent them from the possible harm which comes from the outcome of this manipulation. By claiming that there is no need for this protection, supporters of this objection implicitly adopt either of two positions on citizens' welfare. They either think that people will make choices which are in their best interest regardless of the influence of advertisement, or that governments should not protect their citizens from harm.

We know that the former is not the case, people do not always make decisions which

are best for them (Kahneman 2011; Sunstein and Thaler 2003; Thaler and Sunstein 2003, 2008), especially not when the CAs are purposefully designed to this end (Packard 1957; Tversky 1969; Tversky and Kahneman 1981). Support for the latter is based on an idea that government intervention reduces freedom for both citizens and companies who advertise. Nudges, however, do not reduce nudgees' freedom, as they conserve choice-sets and do not lead them towards a choice against their interest, nor advertisers' freedom since it does not prevent them from advertising. Furthermore, in accordance with Mill's harm principle, power over a member of a society can be "rightfully exercised" when used for "prevent[ing] harm from others" (Mill 1989, 21–2).

Nudges which counter manipulation are not directed towards a specific choice, and only prevent harm from advertisement, and harm prevention is a legitimate reason for governments to counter manipulation. While it is legitimate, it does not, however, mean that they have a duty to do so, and in fact, some might argue that they do not, and that governments should not prevent advertisement because there is a value in it. This argument applies less to nudges than to complete bans on advertisements (which I address next) but I consider how counter-manipulative nudges are compatible with such a view of advertisement.

There might indeed be some value in advertisement. Even though they are manipulative, some people might see them as informative, or discover some products, or brands, they previously did not know about. Some also have artistic and creative value.[11] Some people might also be very good in resisting their manipulative aspect, and simply enjoy seeing advertisements. The exact reason why people might enjoy advertisement is in itself irrelevant since their enjoyment is alone a reason to give them the opportunity to enjoy them.

There are, however, strong reasons to believe that what people might enjoy in advert-

[11] French painter Henri de Toulouse-Lautrec, for example is famous for having started his career making advertisement prints.

isements is not their manipulative aspect. Admitting that manipulation is not always objectionable, most people are likely to prefer to not be manipulated rather than manipulated. This means that a nudge which prevents an advertisement from manipulating its audience is not objectionable from individuals who enjoy advertisements. Counter-manipulative nudges might even make advertisements more enjoyable by removing the fear to be manipulated. Furthermore, even if what someone enjoys in advertisement is its manipulative aspect, the fact that nudges are resistible means that even this aspect can be preserved.

Countering manipulation from advertisement qualifies as preventing harm from others, and it is therefore legitimate for governments to create policies which counter it. Furthermore, the use of nudges allows for the preservation of features of advertisements that people might either enjoy or want to keep. The second objection one might raise against using nudges is that to counter manipulation from advertisement, governments can simply ban advertisement. The next section entertains this objection and argues that nudges can achieve the same results in terms of counter-manipulation, without having the same adverse effects.

3.2 Why not ban advertisement then?

Since advertisements are manipulating us into buying products which negatively impact our welfare, many—including London mayor Sadiq Khan (BBC 2018)—argue in favour of a ban for advertisement for products which are deemed harmful (Gortmaker et al. 2011). The reasoning for that is that less exposure to these products will decrease their consumption which would foster public health and lower spending in that area. While the argument is often made about harmful products, it can be extended to all advertisements for being manipulative. A ban would therefore supposedly avoid manipulation, regardless of the

value (harmfulness or benefit) of the advertised product for specific individuals.

Such bans already exist in some countries, either for some products, or during specific time-frames. In France, all types of advertisement for alcohol and tobacco products have been banned since 1991, and the European Union has tighter rules about fast food advertisement during children's programmes on television. I will not engage with whether these are useful or not, but with the objection towards the use of nudges which would rather see advertisements banned instead. There have not been any comparisons between nudging against and banning manipulative advertisement, so I will use common arguments against advertisement bans and show that they are problematic since they do not solve the issue of manipulation which is of interest here. I conclude that nudges can counter manipulation as well as bans on advertisement, while avoiding issues inherent to bans.

There are three main arguments against advertisement bans to consider. Firstly, banning advertisement restricts the freedom of companies, which can claim that their right to free speech is violated by such a ban. Secondly, one can claim that since there are so many advertisements manipulating us in different directions, they 'cancel each other out' and ultimately do not have an effect on consumers' behaviour. And thirdly, one could argue that banning advertisement would result in preventing small businesses and ventures which either provide free content or rely on advertisement funding to exist from operating.

Whether companies do have a right to free speech through advertisements, and especially whether using this right to manipulate an audience is a legitimate use, are questions for another analysis. While I doubt that many would answer 'yes' to the second question, I will entertain the thought for the sake of the argumentation. Assuming nonetheless that they have this right and are allowed to use it in a manipulative way, a ban would surely infringe upon this right. Using nudges, however, would not. Counter-manipulative nudges would only prevent the audience from being manipulated while still allowing companies to present their products in the way they do today. Although banning advertisement would

potentially lead to a violation on the side of the companies, nudges preserve their freedom.

For the second argument, I cannot evaluate the claim that many advertisement in several different directions reduce their manipulative effect as it is an empirical question. Assuming that it is the case, and assuming that individual advertisements lose their manipulative character as they are offset by other advertisements for the same product, the problem remains as consumers would still be manipulated into wanting the aforementioned product. As an example, let us imagine three brands of biscuits which advertise their products. While consumers would not be manipulated into buying either of these brands' biscuits, they would still be manipulated into buying biscuits. Since even conflicting advertisement messages remain manipulative as they increase consumption, banning advertisement seems to remain attractive. Nudges, however, would prevent the remaining effect, and prevent this increase in biscuit consumption.

The third argument is perhaps the least straightforward because it relates to an issue we tend to think less about. It follows from the idea that some companies or contents would not exist if they could not draw revenues from advertisement. Common examples are free-content internet websites which are only able to maintain themselves by displaying advertisements, or small production films which could not be made without product placements. This objection, however, does not consider that banning advertisement would most likely challenge the organisation of several industries and trigger the emergence of new ways of financing. Nudges, however, would allow such financing to remain while still preventing manipulation, and are therefore more attractive in this case.

I have explored three arguments which object to a complete ban on advertisement, and showed that using nudges can solve these objections while still being effective against manipulation. Nudges seem to bring the best of both worlds as they provide the counter-manipulative benefits and avoid the drawbacks that a ban would incur. Furthermore, as I have discussed in the previous section, nudges preserve the freedom of companies

and advertisement-enthusiasts who both have a preference towards advertisement being available. With regards to harmful products, however, people might be less receptive to anti-ban arguments such as the one about free speech, which might lead policymakers to consider bans such as the ones which are already in place.[12]

If generalised, the use of counter-manipulative nudges against advertisement could draw advertisement away from being manipulative, and closer to being informative. Indeed, if the use of manipulative techniques is made ineffective, advertisers would have to switch to promoting the characteristics of what is advertised in order to rationally convince clients. Such an approach would be welcome since it would be more respectful of people's decision making processes.

[12]While this thesis favours the use of nudges, it is not against bans altogether, and sufficient political support for either in a country is what will ultimately lead to choosing between the two.

Concluding remarks

Despite having received attention from academics and policy makers alike, both the ethics and the possible justifications for the use of nudges still need to be explored in several areas. The goal of this research was to continue this exploration and determine whether governments can legitimately use nudges as counter-manipulation mechanisms against manipulative advertisement. I have argued that it is the case. I have defined nudges emphasising that they are human-made and intentional, rely on cognitive biases, and are easily resistible. Based on this definition, I stipulated characteristics of nudges which are important for their use, such as their alignment with nudgees' preferences, and their level of transparency as well as the accompanying publicity requirements. I also demonstrated that nudges and traditional policies only differ in the way they achieve compliance, and that this difference does not bring further requirements for justification on the side of nudges. This means that the use of nudges is justified whenever the use of traditional policies is.

The fact that nudges do not require more justification than traditional policies does not, however, mean that they can be implemented without further considerations. Discussing nudges, I have established that their resistibility is a very important feature, and that a strong focus on their transparency is required to ensure it. I separated nudges into visible and transparent, depending on whether they are either noticeable without their goals being easily identifiable, or noticeable with easily identifiable goals. While transparent nudges whose goals are easily identifiable can be implemented as such, visible nudges whose goals are unclear for the nudgee require further publicity.

Provided that they satisfy these requirements, I identified two ways nudges can be used to counter manipulation through advertisements. First, they can modify citizens' CAs further by adjusting the salience of information away from the manipulative messages delivered by the advertisers. Second, they can simply modify nudgees' CAs regardless of previous

advertisement, and arrange them in a way which fits citizens' goals and values. Considering two possible objections to the argument, I justified the need for counter-manipulation explaining that it qualifies as prevention from harm, which justifies an intervention from the government. I further rejected the option to ban advertisements on the grounds that although it would effectively prevent manipulation through advertisement, it has negative effects which would be avoided by using nudges.

By suggesting that nudges can be used as counter-manipulation mechanisms against advertisement, this thesis expands the possible uses of nudges. Counter-manipulative nudges respect, and even increase, individuals' autonomy by making sure that they do not make decisions which are not in line with their preferences. The possibility to apply nudges as counter-manipulation mechanisms expands governments' toolkit in terms of consumer protection.

Furthermore, the idea that nudges do not require more justification than traditional policies is not specific to their use as a counter-manipulation mechanism and can be applied in other areas. While many scholars address nudges and study their use, or specific characteristics, few scholars have questioned whether the justificatory bar is higher for nudges than for traditional policies. More research is needed in this area, in order to determine whether there are cases in which the application of nudges does requires more justification than traditional policies. Exploring further cases, especially ones in which a slight change in behaviour can have drastic impacts on individuals' lives, could improve the design of further policies, by taking behavioural insights into account, and incorporating nudges within broader policy frameworks.

Bibliography

Alemanno, Alberto, and Anne-Lise Sibony. 2015. *Nudge and the Law A European Perspective*. Ed. by Alberto Alemanno and Anne-Lise Sibony. London: Hart Publishing.

Alemanno, Alberto, and Alessandro Spina. 2014. "Nudging Legally: on the Checks and Balances of Behavioural Regulation". *International Journal of Constitutional Law* 12 (2).

Allais, Maurice. 1953. "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine". *Econometrica* 21 (4): 503–546.

Ariely, Dan. 2008. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. Harper Collins Publishers.

BBC, News. 2018. "London Mayor Sadiq Khan plans TfL 'junk food' advert ban". http://www.bbc.com/news/uk-england-london-44071809.

Beauchamp, Tom L., R. M. Hare and Barry Biederman. 1984. "Manipulative Advertising". *Business & Professional Ethics Journal* 3 (3): 1–30.

Benartzi, Shlomo, et al. 2017. "Should Governments Invest More in Nudging?" *Psychological Science* 28 (8): 1041–1055.

Blumenthal-Barby, J. S. 2013. "Choice Architecture: Improving Choice While Preserving Liberty?" In *Paternalism*, ed. by Christian Coons and Michael Weber. Cambridge University Press.

Blumenthal-Barby, J. S., and Hadley Burroughs. 2012. "Seeking Better Health Care Outcomes: The Ethics of Using the "Nudge"". *The American Journal of Bioethics* 12 (2): 1–10.

Blumenthal-Barby, J. S., et al. 2013. "Methods of Influencing the Decisions of Psychiatric Patients: An Ethical Analysis". *Harvard Review of Psychiatry* 21 (5): 276–279.

Bovens, Luc. 2008. "The Ethics of Nudge". In *Preference Change: Approaches from Philosophy, Economics and Psychology*, **by** Till Grüne-Yanoff and S.O. Hansson. Berlin and New York: Springer.

Bowles, Samuel. 2008. "Policies Designed for Self-Interested Citizens May Undermine "The Moral Sentiments": Evidence from Economic Experiments". *Science* 320:1605–1609.

Bowles, Samuel, and Sung-Ha Hwang. 2008. "Social preferences and public economics: Mechanism design when social preferences depend on incentives". *Journal of Public Economics* 92 (8): 1811–1820.

Bruce, Norris I., B.P.S. Murthi and Ram C. Rao. 2017. "A Dynamic Model for Digital Advertising: The Effects of Creative Format, Message Content, and Targeting on Engagement". *Journal of Marketing Research* 54 (2): 202–218.

Bullock, Emma C. 2014. "Free Choice and Patient Best Interests". *Health Care Analysis* 24 (4): 374–392.

Burgess, Adam. 2012. "'Nudging' Healthy Lifestyles: The UK Experiments with the Behavioural Alternative to Regulation and the Market". *European Journal of Risk Regulation* 3 (1): 3–16.

Chetty, Raj. 2015. "Behavioral Economics and Public Policy: A Pragmatic Perspective". *American Economic Review Papers and Proceedings* 105 (5): 1–33.

Cohen, Shlomo. 2013. "Nudging and Informed Consent". *The American Journal of Bioethics* 13 (6): 3–11.

Congdon, William J., Jeffrey R. Kling and Sendhil Mullainathan. 2011. *Policy and Choice: Public Finance through the Lens of Behavioral Economics*. Washington: Brookings Institution Press.

Coons, Christian, and Michael Weber. 2014. *Manipulation: theory and practice*. Ed. by Christian Coons and Michael Weber. Oxford ; New York: Oxford University Press.

Damgaard, Mette Trier, and Christina Gravert. 2016. "The Hidden Costs of Nudging: Experimental Evidence from Reminders in Fundraising".

Danciu, Victor. 2014. "Manipulative marketing: persuasion and manipulation of the consumer through advertising". *Theoretical and Applied Economics* 21 (2): 19–34.

Dolan, Paul, et al. 2010. *MINDSPACE: Influencing behaviour through public policy*.

Dworkin, Gerald. 2013. "Lying and nudging". *Journal of Medical Ethics* 39 (8): 496–497.

Economist, The. 2017. "Policymakers around the world are embracing behavioural science". https://www.economist.com/news/international/21722163-experimental-iterative-data-driven-approach-gaining-ground-policymakers-around.

Escaron, Anne L., et al. 2013. "Supermarket and Grocery Store–Based Interventions to Promote Healthful Food Choices and Eating Practices: A Systematic Review". *Preventing Chronic Disease* 10:1–20.

Friedman, Milton. 1966. "Essays In Positive Economics". In *The Methodology of Positive Economics*, 3–43. Chicago: University of Chicago Press.

Gigerenzer, Gerd. 2015. "On the Supposed Evidence for Libertarian Paternalism". *Review of Philosophy and Psychology* 6 (3): 361–383.

Glod, William. 2015. "How Nudges Often Fail to Treat People According to Their Own Preferences". *Social Theory and Practice* 41 (4): 599–617.

Goldstein, Noah J., Robert B. Cialdini and Vladas Griskevicius. 2008. "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels". *Journal of Consumer Research* 35 (3): 472–482.

Gortmaker, Steven L., et al. 2011. "Changing the Future of Obesity: Science, Policy and Action". *Lancet* 378 (9793): 838–847.

Grand, Julian Le, and Bill New. 2015. *Government Paternalism: Nanny State or Helpful Friend?* Princeton University Street.

Grüne-Yanoff, Till. 2012. "Old wine in new casks: libertarian paternalism still violates liberal principles". *Social Choice and Welfare* 38 (4): 635–645.

Grüne-Yanoff, Till, and Ralph Hertwig. 2015. "Nudge Versus Boost: How Coherent are Policy and Theory?" *Minds and Machines* 26:149–183.

Guala, Francesco, and Luigi Mittone. 2015. "A Political Justification of Nudging". *Review of Philosophy and Psychology* 6 (3): 385–395.

Hallsworth, Michael. 2017. "Rethinking public health using behavioural science". *Nature Human Behaviour* 1 (9): 612–612.

Hansen, Pelle Guldborg, and Andreas Maaløe Jespersen. 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.

Hansen, Pelle Guldborg, Laurits Rohden Skov and Katrine Lund Skov. 2016. "Making Healthy Choices Easier: Regulation versus Nudging". *Annual Review of Public Health* 37 (1): 237–251.

Haushofer, J., and E. Fehr. 2014. "On the psychology of poverty". *Science* 344 (6186): 862–867.

Hausman, Daniel M., and Brynn Welch. 2010. "Debate: To Nudge or Not to Nudge". *Journal of Political Philosophy* 18 (1): 123–136.

Hertwig, Ralph, and Till Grüne-Yanoff. 2017. "Nudging and Boosting: Steering or Empowering Good Decisions". *Perspectives on Psychological Science* 12 (6): 973–986.

Hill, Claire. 2007. "Anti-Anti-Anti-Paternalism". *NYU Journal of Law & Liberty* 2 (444): 444–454.

John, Peter, Michael Sanders and Jennifer Wang. 2014. "The Use of Descriptive Norms in Public Administration: A Panacea for Improving Citizen Behaviours?" *SSRN Electronic Journal*.

Johnson, Eric J., and Daniel Goldstein. 2003. "Do Defaults Save Lives?" *Science* 302 (5649): 1338–1339.

Johnson, Eric J., et al. 2013. "Can Consumers Make Affordable Care Affordable? The Value of Choice Architecture". *PLOS ONE* 8 (12): 1–6.

Kahneman, Daniel. 2003. "A perspective on judgment and choice: mapping bounded rationality". *American Psychologist* 58 (9): 697–720.

— . 2011. *Thinking, Fast and Slow*. London: Penguin Books.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk". *Econometrica* 47 (2): 263–292.

— . 1996. "On the Reality of Cognitive Illusions". *Psychological Review* 130 (3): 582–591.

Kraft-Todd, Gordon, et al. 2015. "Promoting Cooperation in the Field". *Current Opinion in Behavioral Sciences* 3:96–101.

Krawczyk, Michal, and Fabrice Le Lec. 2015. "Can we neutralize social preference in experimental games?" *Journal of Economic Behavior & Organization* 117:340–355.

Lacetera, Nicola, and Mario Macis. 2009. "Do All Material Incentives for Prosocial Activities Backfire? The Response to Cash and Non-Cash Incentives for Blood Donations".

Lodge, Martin, and Kai Wegrich. 2016. "The Rationality Paradox of Nudge: Rational Tools of Government in a World of Bounded Rationality: The Rationality Paradox of Nudge". *Law & Policy* 38 (3): 250–267.

Loewenstein, George. 2011. "Confronting reality: pitfalls of calorie posting". *The American Journal of Clinical Nutrition* 93 (4): 679–680.

Loewenstein, George, et al. 2012. "Can behavioural economics make us healthier?" *BMJ* 344 (1): 3482–3485.

Loewenstein, George, et al. 2015. "Warning: You Are About to be Nudged". *Behavioral Science & Policy* 1 (1): 35–42.

Madrian, Brigite C., and Dennis F. Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behaviour". *The Quarterly Journal of Economics* 116 (4): 1149–1187.

Main, Marisa E. 2012. "Simply Irresistible: Neuromarketing and the Commercial Speech Doctrine". *Duquesne Law Review* 50:605–643.

Mathios, Alan D. 2000. "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market". *The Journal of Law & Economics* 43 (2): 651–678.

Mill, John Stuart. 1989. *On Liberty*. Cambridge: Cambridge University Press.

Mitchell, Gregory. 2005. "Libertarian Paternalism is an Oxymoron". *Northwestern University Law Review* 99 (3): 1245–1277.

Moles, Andrés. 2015. "Nudging for Liberals". *Social Theory and Practice* 41 (4): 644–667.

Nagatsu, Michiru. 2015. "Social Nudges: Their Mechanisms and Justification". *Review of Philosophy and Psychology* 6 (3): 481–494.

Noar, Seth M., et al. 2016. "Pictorial cigarette pack warnings: a meta-analysis of experimental studies". *Tobacco Control* 25:341–354.

Noggle, Robert. 1996. "Manipulative Actions: A Conceptual and Moral Analysis". *American Philosophical Quarterly* 33 (1): 43–55.

— . 2018. "Manipulation, salience, and nudges". *Bioethics* 32 (3): 164–170.

Nys, Thomas RV, and Bart Engelen. 2017. "Judging Nudging: Answering the Manipulation Objection". *Political Studies* 65 (1): 199–214.

O'Dowd, Adrian. 2017. "Spending on junk food advertising is nearly 30 times what government spends on promoting healthy eating". *BMJ - British Medical Journal* 359:j4677.

Ohman, Arne, Anders Flykt and Francisco Esteves. 2001. "Emotion Drives Attention: Detecting the Snake in the Grass". *Journal of Experimental Psychology* 130 (3): 466–478.

Packard, Vance. 1957. *The hidden persuaders*. New York: D. McKay Co.

Papies, Esther K., et al. 2013. "Using health primes to reduce unhealthy snack purchases among overweight consumers in a grocery store". *International Journal of Obesity* 38 (4): 597–602.

Phillips, Michael J. 1997. *Ethics and Manipulation in Advertising: Answering a Flawed Indictment*. Greenwood Publishing Group.

Quigley, Muireann. 2013. "Nudging for health: on public policy and designing choice architecture". *Medical Law Review* 21 (4): 588–621.

Rayner, G., and T. Lang. 2011. "Is nudge an effective public health strategy to tackle obesity? No". *BMJ - British Medical Journal* 342.

Reisch, Lucia A., and Cass Sunstein. 2016. "Do Europeans Like Nudges?" *SSRN Electronic Journal*.

Rizzo, Mario J., Douglas Glen and Whitman. 2009. "The Knowledge Problem of New Paternalism". *Brigham Young University Law Review*.

Roberto, Christina A., and Ichiro Kawachi. 2015. *Behavioral Economics and Public Health*. Oxford: Oxford University Press.

Roberto, Christina A., and Neha Khandpur. 2014. "Improving the design of nutrition labels to promote healthier food choices and reasonable portion sizes". *International Journal of Obesity* 38:S25–S33.

Saghai, Yashar. 2013. "Salvaging the concept of nudge". *Journal of Medical Ethics* 39 (8): 487–493.

Sandroni, Alvaro, and Francesco Squintani. 2007. "Overconfidence, Insurance, and Paternalism". *The American Economic Review* 97 (5): 1994–2004.

Schwartz, Janet, et al. 2012. "Inviting Consumers To Downsize Fast-Food Portions Significantly Reduces Calorie Consumption". *Health Affairs* 31 (2): 399–407.

Selley, Eliza, and Michael Sanders. 2017. "Inspirational students encourage university applications". http://www.behaviouralinsights.co.uk/uncategorized/inspirational-students-encourage-university-applications/.

Shabbir, Haseeb, and Des Thwaites. 2007. "The Use of Humor to Mask Deceptive Advertising: It's No Laughing Matter". *Journal of Advertising* 36 (2): 75–85.

Shah, Anuj K., Sendhil Mullainathan and Eldar Shafir. 2012. "Some Consequences of Having Too Little". *Science* 338 (6107): 682–685.

Shampanier, Kristina, Nina Mazar and Dan Ariely. 2007. "Zero as a Special Price: The True Value of Free Products". *Marketing Science* 26 (6): 742–757.

Sibony, Anne-Lise, and Geneviève Helleringer. 2015. "EU Consumer Protection and Behavioural Sciences: Revolution or Reform?" In *Nudge and the Law: A European Perspective*, ed. by Alberto Alemanno and Anne-Lise Sibony, 209–233. Bloomsbury Publishing.

Simon, Herbert A. 1947. *Administrative Behavior: a Study of Decision-Making Processes in Administrative Organization*. New York City, NY: Macmillan Publishers.

— . 1955. "A Behavioral Model of Rational Choice". *The Quarterly Journal of Economics* 69 (1): 99–118.

— . 1985. "Human Nature in Politics: The Dialogue of Psychology with Political Science". *The American Political Science Review* 79 (12): 293–304.

Skov, Laurits Rohden, et al. 2013. "Choice architecture as a means to change eating behaviour in self-service settings: a systematic review". *Obesity Reviews* 14 (3): 187–196.

Smith, Gerald E. 1996. "Framing in advertising and the moderating impact of consumer education". *Journal of Advertising Research* 36 (5): 49–64.

Stallen, Mirre, and Alan G. Sanfey. 2015. "Cooperation in the brain: neuroscientific contributions to theory and policy". *Current Opinion in Behavioral Sciences* 3:117–121.

Sunstein, Cass. 1988. "Beyond the Republican Revival". *The Yale Law Journal* 97 (8): 1539.

— . 2011. "Empirically Informed Regulation". *University of Chicago Law Review* 78 (4): 1349–1429.

— . 2014. "Nudging: A Very Short Guide". *J. Consumer Pol'y* 37 (583).

— . 2015a. "Do People Like Nudges?" *SSRN Electronic Journal*.

— . 2015b. "Nudging and Choice Architecture: Ethical and Considerations". *Yale Journal on Regulation* 32 (2): 1–50.

— . 2017a. *Human Agency and Behavioral Economics: Nudging Fast and Slow*. New York City, NY: Palgrave Macmilan.

— . 2017b. "People Like Nudges (Mostly)". In *Human Agency and Behavioral Economics: Nudging Fast and Slow*, 17–39. New York City, NY: Palgrave Macmilan.

Sunstein, Cass, Lucia A. Reisch and Julius Rauber. 2017. "Behavioral Insights All Over the World? Public Attitudes Toward Nudging in a Multi-Country Study". *SSRN Electronic Journal*.

Sunstein, Cass, and Richard Thaler. 2003. "Libertarian Paternalism Is Not an Oxymoron". *The University of Chicago Law Review* 70 (4): 1159–1202.

Team, Behavioural Insights. 2010. *Applying behavioural insight to health*. Report. Behavioural Insights Team.

Thaler, Richard. 1980. "Toward a positive theory of consumer choice". *Journal of Economic Behavior & Organization* 1 (1): 39–60.

— . 2010. *Rethinking Regulation after the Financial Crisis and the Oil Spill: A Behavioral Approach: https://www.youtube.com/watch?v=zLMephdISTw.*

— . 2015. *Misbehaving: The Making of Behavioral Economics.* New York City, NY: W. W. Norton & Company.

Thaler, Richard, and Cass Sunstein. 2003. "Libertarian Paternalism". *The American Economic Review* 93 (2): 175–179.

— . 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Yale University Press.

Thaler, Richard, Cass Sunstein and John Balz. 2012. "Choice Architecture". In *The Behavioral Foundations of Public Policy*, ed. by Eldar Shafir. Princeton University Press.

Treise, Debbie, et al. 1994. "Ethics in Advertising: Ideological Correlates of Consumer Perceptions". *Journal of Advertising* 23 (3): 59–69.

Tversky, Amos. 1969. "Intransitivity of Preferences". *Psychological Review* 76 (1): 30–48.

Tversky, Amos, and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability". *Cognitive Psychology* 5:207–232.

— . 1974. "Judgment under Uncertainty: Heuristics and Biases". *Science* 185 (4157): 1124–1131.

— . 1981. "The framing of decisions and the psychology of choice". *Science* 211 (4481): 453–458.

— . 1983. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment". *Psychological Review* 90 (3): 293–315.

Waldron, Jeremy. 2014. "It's All for Your Own Good". http://www.nybooks.com/articles/2014/10/09/cass-sunstein-its-all-your-own-good/.

Wänke, Michaela. 2016. "Primes as hidden persuaders". *Current Opinion in Psychology* 12:63–66.

Wilkinson, T. M. 2013. "Nudging and Manipulation". *Political Studies* 61 (2): 341–355.

— . 2017. "Counter-Manipulation and Health Promotion". *Public Health Ethics* 10 (3): 257–266.

Wisdom, Jessica, Julie S. Downs and George Loewenstein. 2010. "Promoting Healthy Choices: Information versus Convenience". *American Economic Journal: Applied Economics* 2 (2): 164–178.

Zhang, Hongxia, et al. 2014. "Be rational or be emotional: advertising appeals, service types and consumer responses". *European Journal of Marketing* 48 (11): 2105–2126.