

On the Concept of Nudging: Its Coherence and Applicability?

Mads Wistrup Freund

September 15, 2017

25-June-1993-xxxx

MSc in Business Administration and Philosophy

Vejleder: Marius Gudman Høyer



Abstract

The practice of nudging has attracted attention from researchers, policymakers, and practitioners alike. The contemporary interest in nudging has sparked demands for a clearer examination of the conditions under which nudging and other behavioral methods can be used efficiently and acceptably. There are several unresolved questions which may have contributed to heated political and scientific discussions, especially heated is the ethical discussion of applying nudging in the public policy domain. This thesis concentrate on the philosophical premises nudges are built on. It is argued that the model is philosophically problematic. Different interventions might operate through various logics that threaten not only effectiveness but also representativeness and ethical ends. This thesis argues that the next step in advancing the acceptability nudging requires a clarification. The thesis investigates the internal logic of this practice by exploring the expositions and implications within the literature on nudging in public policy, by specifically exploring its origins in behavioral economics; its impact on welfare policy; and ethical implications. This analysis is guided by Foucault's concept of veridiction and understanding of power. This thesis concludes that nudging would be more acceptable if it clarified that humans, in these models, are treated as defective econs. Choice architects are trying to represent the complex reality of human judgment and decision-making with the inclusion of psychology in a highly simplified normative modeling framework, to increase individuals tendency to act according to neo-classical rational choice models. Advancing research in nudging is necessary to distill the value of nudging interventions and equally important to provide tangible cases to debates such as ethical appropriateness, effectiveness, and public approval.

Contents

1 Chapter. Exposition of Propositions in Nudging and Critique Thereof	15
1.1 Political Ideal of Libertarian Paternalism	15
1.2 System 1 and System 2	16
1.3 The Libertarian Critique	17
1.4 The Anti-Libertarian Critique	19
1.5 Common Reply by Thaler and Sunstein	20
2 Chapter. Exposition of Dual process Theory and Decision-Making	22
2.1 Starting Assumptions Bounded Rationality	22
2.2 Human and Econs – Thaler and Sunstein’s Application of Dual-Process Theory	24
2.3 How to Identify Human Biases	25
2.4 The Influence Exerted by Default Choices	26
2.5 The Power of Framing	26
2.6 Social Biases	27
2.7 Habits and their Influence on Choice	29
2.8 Reinforcing Self-Control Strategies	30
2.9 Rationality	30
3 Chapter 3. Welfare Economics – Combining Nudging and Preferences	32
3.1 Neoclassical Economics	32
3.2 Neoclassical Economic Framework for Implementing Public Policy - Game Theory.	33
3.3 7.1 Homo Economicus and Humans	33
3.4 Behavioral Economics – Explaining Deviations from Neoclassical Economics	34
3.5 Three Approaches to Behavioral Change	35
3.5.1 1. Provision of Information - the ‘Think’ Approach	35
3.5.2 2. Direct Regulation - the Shove Approach	35
3.5.3 3. Choice architecture - the Nudge Approach	36
3.6 Why Nudging is an Attractive Tool in Public Policy	37
3.7 Measuring Welfare Improvements from Nudging	37
4 Chapter. Definition and Categorization of Nudging	40
4.1 Significantly Changing Incentives	41
4.2 Provision of Information	43
4.3 Health Policy Example	44
4.4 Revisiting the Anti-Nudge Position	45
4.5 As Judged by Themselves	46
4.6 Preferences	47
5 Chapter. Intentions and Ethics	51
5.1 Manipulation	51
5.2 Evaluating Nudging	51
5.3 Transparency	52
5.4 Nudging an Intentional Act	54
5.5 Manipulation and its Context	54
5.6 Blood Donation	55
5.7 Self-Interested Nudges	55
5.8 Responsiveness to Nudging	56
5.9 Can Nudging be Resisted?	57
6 Chapter. Challenging implications as to scientific findings and implementation in public policy	59
6.1 Welfare	59
6.1.1 Lasting?	59
6.1.2 Practical Evidence	60
6.2 Implementation Problems	61
6.2.1 A Taxonomy of Nudges	61
6.2.2 Public Permission	62
6.2.3 Mandated choice or presumed consent	63
6.2.4 Dubious Nudging	64
6.2.5 Political Critique	65
6.2.6 The Current Status of Nudging	65
6.3 Conclusion	66
7 Bibliography	71
References	71

Introduction

Nudging is a tool which is increasingly being applied in public policy OECD (2017). Nudging is applied by policymakers, private companies, and practitioners alike OECD (2017). Government-backed organizations are emerging to advise politicians about the use of nudges in both Anglo-Saxon countries and several other countries in Europe Blumenthal-Barby and Burroughs (2012); Whitehead, Jones, Howell, Lilley, and Pykett (2014). With its promise of influencing behavior without changing our incentives nor forbidding any actions it is evident why the interest in nudging has spiked. The publications of books from prominent scholars increased the awareness of these models. Two of the most influential is the collaboration between the American legal scholar Cass Robert Sunstein 1954 and the American economist Richard H. Thaler 1945's book *Nudge - Improving health, wealth, and happiness* (2008) and the Israeli-American psychologist Daniel Kahneman 1934's book *Thinking, fast and slow* (2011). These publications, explains how decision-making contexts may on a regular basis lead individuals to act contrary to their well-informed intentions. The publications assert that people base most judgments and choices on routines and not on deliberative thought. As a result, minor changes to the context in which people are making decisions - 'choice architecture' – can have substantial effects on behavior. The application of behavioral economics to policy is primarily used in certain policy areas, such as pensions, tax, sustainability and consumer protection C. R. Sunstein (2016b). An example is the European Commission's ban on pre-ticked boxes in online sales, based on evidence that default settings can have a powerful influence on choice Ly and Soman (2013).

The central idea of nudging is that small and apparently insignificant details, such as the pre-ticked boxes, can have a major impact on people's behavior. Nudging is a low-cost intervention that can influence our everyday choices and behaviors without recourse to prohibition or changes to incentives. Instead of facing the threat of reactance through bans or taxes to discourage consumption of junk food, people can be nudged into healthier behavior, for instance by placing healthy food in a prominent position. In contrast to formerly existent and more direct state interventions, nudging does not simply aim to impose costs to produce a specific behavior. Nudging intends to influence thought processes and human rationality. Nudging is an intervention that policymakers and others with public responsibilities for a group of people can implement to help individuals with their decisions Sunstein and Thaler (2008). Past examples of using nudging to tackle societal issues are organ donation, obesity, retirement savings, traffic, and environmental issues (Hollands et al., 2015; Sunstein & Thaler, 2008; Whitehead et al., 2014).

In its ideal form nudging seemingly provides public policymakers with the ultimate tool: An ethical and politically uncontroversial approach to influencing the choices and behavior of citizens in accordance with their interests. The elusiveness of nudging has raised some concerns

as to whether nudging can be ethically justified because it is perceived as a slippery slope to manipulate people without their consent or knowledge of it directly. The British finance professor Ricardo Rebonato perceives nudging as nothing short of manipulation. It is worse than paternalism because it removes from citizens their claimed freedom to choose Rebonato (2014). According to Ricardo, governments cannot apply nudging in policy interventions, as it will lose its legitimacy as a democratic institution. Supporters of this critique argue that the mere disclosure of nudging might decrease the effectiveness of the nudge.

The British Philosopher Luc Bovens claims: “*nudges work better in the dark*” (Bovens, 2009, p.3). He holds this view precisely because the nudges interference with the recipient is more or less outside of conscious awareness. If nudging or some sub category of nudging manipulates people, then these instances would be incompatible with public policymaking in a modern democracy Hansen and Jespersen (2013). The statement from (Bovens, 2009) cited above put the attention to why it is important to review empirical findings. To study whether disclosure has an effect on the effectiveness of nudges and under what circumstances. Furthermore, the thesis questions whether nudging - and the assumptions it is based on - is appropriate to public policy. If people are nudged towards making improved decision-making, we must first investigate how we understand decision-making to distill the value of nudging interventions. The use of behavioral economics can to some extent cause fundamental problems for conventional economics by calling revealed preferences into question. The responses to this issue are critically examined, such as the reply recently endorsed by Hausman (2011, :102). This thesis will present nudging as build on the idea of ‘preference purification.’ This idea proposes a method by which individual decisions, inconsistent with the logical assumptions in rational choice, can be treated as mistakes. By treating these choices as mistakes, behavioral economics are then able to reconstruct the preferences that the individual would have acted upon, had their reasoning not been distorted by whatever psychological mechanisms responsible for the mistake. The task is to use the satisfaction of these reconstructed preferences as a normative criterion for policy. This thesis will argue that nudging is neither free of manipulative aspects nor can it be characterized as simple manipulation. It is concluded, that, if used properly nudging can, in fact, improve well-being.

The Problem

Welfare economics is about coordinating individuals not, ultimately, the individuals themselves. Psychology focuses on helping individuals as individuals, to overcome irrationality. We need some assumption about systematic behavior, to understand individuals’ interaction within a system. Most importantly is bounded selective rationality a useful assumption for policy-makers? This thesis will focus on nudging and its applicability as a tool for behavior change in general. The focus of this research aims at understanding the translation of the behaviorally informed

tool nudging into welfare policy. This research opts to concentrate on the arguments for nudges, thus refraining from analyzing the specific outcome of nudge interventions. This research also aims at elaborating predominant assumptions and opinions within the discourse of nudging. In this thesis, Thaler and Sunstein's book from 2008 *Nudge: Improving decisions about health, wealth, and happiness* and other selected sources is analyzed concerning the exercise of influence via nudging, the topic of government intervention and the possible hegemony of nudging.

Since Thaler and Sunstein's publication in 2008, a plethora of interventions has been labeled nudging. The confusion is partly attributable to Thaler and Sunstein's broad definition of nudging, which have caused conceptual issues making the label confusing. The contemporary use of nudging has become a favorite name for describing almost any type of intervention based on behavioral insights, and a more restricted definition is, therefore, employed in this thesis. Instead of asking whether one can accept or reject nudging, this thesis asks how the nudge literature perceive rationality, preferences, decisions, and choices. First this thesis asks, under what conditions do nudges work like they are supposed to work, and the second asks whether these nudges can be endorsed normatively. It is assumed that clarifying these assumptions illuminates the nudge concept, by providing clarity in this regard. It seems clear that the normative problems related to nudging cannot be divided into 'transparent' on the one hand and 'non-transparent.' It is believed that clarifying this continuum will improve the usefulness of nudging. This paper does not seek to expand on a further line of critical engagement with behavioral economics, namely understandings of rationality and assumptions regarding human behavior. As noted by (Rostain, 2000), behavioral economics appeals to those who believe in 'homo economicus' and those who point to the importance of social norms. This thesis aims at contributing to the literature with clarification and coherence to support further debate.

Problem Statement

Can the suspicion of manipulation in relation to nudge interventions be resolved, or is this type of intervention simply too interfering to be compatible in public policy?

Research Question

1. What kind of problem is solved in contexts in which nudges may be preferable to more controlling influences?
2. What kinds of assumptions are needed to endorse nudges completely, both to normatively endorse the assumptions of successful nudges?
3. How do behavioral techniques exercise influence over individuals, can its influence be determined as manipulation, empowerment or both?
4. Can we understand how to coordinate an economic system based on an assumption of predictable irrationality?
5. Can nudges be a part of Pareto-efficient economic policies? Alternatively, are they helping individuals to choose and decide more efficiently, in terms of individual welfare regardless of the wider society?

Literature Review

The articles and books selected for the literature review categorize as academic replies, where the authors evaluate Thaler and Sunstein's concepts of nudging and libertarian paternalism or articles and books debating the assumptions behind dual process theory. Most of the sources in this thesis were published in reaction to Thaler and Sunstein's 2008 publication. The main focus of this thesis is to provide an expanded analysis of the literature concerning or related to nudging. The ideas behind nudging are built on a large complex of intertwined assumptions which will subsequently be disentangled to provide clear picture of the implications and possibilities of applying nudging:

“A nudge, as we will use the term, is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting the fruit at eye level counts as a nudge. Banning junk food does not.” Sunstein and Thaler (2008, p.6).

The notion of nudging is based on three broad claims by Thaler and Sunstein, which can be summarized as follows:

1. Libertarian paternalism is not an oxymoron.
2. Choice architecture, or how options are presented to a decision-maker, has a significant impact on both the decisional process and the final choice.
3. Choice architecture is inevitable: since options have to be presented one way or another, and it makes sense to present options in the best interests of the decision maker.

The theoretical foundations of nudging are located in the compound of the behavioral economics literature on human fallibility. These include heuristics, biases, rules of thumb (Kahneman, Knetsch, & Thaler, 1991; Kahneman & Tversky, 1984) and in libertarian paternalism (Hollands et al., 2015; C. R. Sunstein & Thaler, 2003; Thaler & Benartzi, 2004). (C. R. Sunstein & Thaler, 2003) describes a policy as paternalistic when it is chosen with the intention of influencing the choices of individuals in a way that will improve their status quo objectively and measurably.

The addition liberal paternalism is a political thought that advocates the preservation of freedom of choice, but at the same time authorize private and public institutions to steer people in directions that will promote their welfare. The objective is to avoid, as much as possible, effects of suboptimal decisions without any coercion. (Sunstein & Thaler, 2008) find nudging to be a necessary tool in welfare policy because much of human choice and behavior is performed on autopilot, mostly drawing on mental shortcuts learned throughout previous experiences. Nudges act on that gap by setting up a choice architecture that will guide one's decision making in the direction that the individual him/herself find optimal. In theory, nudges should always maintain freedom of choice, never be manipulative, and lead to protecting people against economic and physical harm.

One of the most cited examples in the literature on libertarian paternalism is the 'Save More Tomorrow' or 'SMarT program'. (Thaler & Benartzi, 2004, pp.179-182) propose a four-step approach where individuals commit in advance to allocate a portion of their future salary increases towards a retirement savings plan. Many economists still believe that libertarian paternalism is an oxymoron and should simply be called paternalism. Their main argument is that if individuals are incapable of deciding what is best for them, nudging just imposes a choice without any real freedom of opting for any of the available alternatives (Rebonato, 2014). That is in practice nudging simply shifts the selection from one option to another. No choice is entailed because the agent does not reflect on the options and nudging is therefore straightforward paternalism.

There are several gaps in understanding human behavior and a lack of understanding aspects of the automatic system. In this thesis, the most noticeable is a lack of comparative research

into the limits of the transferability of behavior change interventions across cultural differences and a lack of understanding about the effect of social dynamics on behavior.

“The majority of experimental evidence about behavior change relates to individual approaches and comes largely from disciplines within psychology(. . .) Much of the evidence is limited, and it is rare that evidence can be extrapolated or generalized from those interventions to the wider population with confidence and without caveats” (Science, Committee, et al., 2011, p.18).

People have also alluded to a lack of proof about the cost-effectiveness of interventions and to a disappointing lack of long-term data to judge the effectiveness of interventions over sustained periods. The government may often wish to establish healthier behavioral patterns. Goals that often requires the individual to struggle for its achievement. Nudging claims to reduce this burden significantly The challenges perceived by the literature includes that nudges can be considered a form of manipulation. There is a lack of consensus in the field regarding whether nudges should be transparent, and that nudging is a buzzword that risks fizzling out because it lacks significant impact.

Method - Applying Foucault to Identify the Complex Assemblage of Heterogeneous Practices

This Master thesis is aimed towards different stakeholders who are involved in implementing nudging or considering it. This study will draw on knowledge in the literature by mapping the various concerns and possibilities from notable authors within the behavioral literature. In general, this thesis aims at analyzing the nudge concept and its corresponding literature inspired by a Foucauldian approach to social analysis.

The literature on nudging oscillates between very different perceptions of human autonomy. As a result, some discussions have an irreconcilable touch. Social scientists such as economists and psychologists study humans, their actions, and the reasons, causes, and consequences of their actions. In studying individual actions, social scientists construct models. Models such as dual process theory help to justify nudging. This thesis deploys the thought of the French philosopher Michel Foucault (1926-1984). Foucault’s work on micro-power and how practices evolve within scientific disciplines are incorporated in exploring the interpretations of the influence nudging exerts over its audience. Another important part of this thesis is the analysis of preferences and the role they play in behavioral economics understanding of human action. By drawing on Foucault’s works and methods, this thesis will question whether the version of rational choice theory deployed by economists and the accounts delivered by behavioral economists are more alike than behavioral economists seems to admit.

Foucault's Conception of Power and Regimes of Practices

Foucault's work comprises of a multiplicity of ideas regarding the often subliminal operation of power, the rationality of the state and the logic of government action towards a population relatable to the concept of nudges. Foucault's concept of 'discourse' may offer both insights regarding the mode of operation of nudges and for a meta assessment of prevailing opinions within the academic sphere surrounding nudging. As the extent of Foucault's work on governmentality, the state and power are enormous. The presented overview will outline only the most relevant concepts from his publications. Furthermore, he exerts, like other philosophers a certain fondness of gnomic aphorisms, which requires a significant amount of examination to understand. Foucault's concern in much of his work on power is to pay close attention to apparently inconsequential practices, to reveal how practices are shaped and made normal. Furthermore, how they interact with other practices, as in (Foucault, 2012) analysis in *Discipline and Punish*.

One often finds oneself in an awkward position when trying to explain Foucault's methodology, especially because he rarely spoke of methodology himself. For Foucault power is neither repressive nor prohibitive. Power is not even opposed to freedom, on the contrary: "*power is exercised only over free subjects, and only insofar as they are free*" (Foucault, 1982, p.790). Returning to one of his classics, *Discipline and Punish* (Foucault, 2012) illustrates an aspect of his method well.

On the one hand, Foucault aims in this book thoroughly to dissect the prison, to break it down into the smallest of its constituent parts, so that we no longer see it as a unified 'thing' but rather as a complex assemblage of heterogeneous practices. The point of this move is to undo any sense of the givenness of the prison, to expose the cracks in the edifice, as it were, and to induce a new sensitivity to the sheer contingency of the intersections that enable its many parts to appear as forming a quasi-natural whole.

The second move in his analysis consists in rebuilding the prison based on the insights from the initial step of the analysis. The process is initiated to generate a new understanding of how the prison has come to existence. The second movement is perhaps the most important to Foucault's project, he states in *Discipline and Punish*:

"The further one breaks down the processes under analysis, the more one is enabled and indeed obliged to construct their external relations of intelligibility" (Foucault, 1991, :77).

In Foucault's view, practices follow their logic; the social order does, as a result, have a degree of coherence and patterning. Foucault state that practices "*possess up to a point their own specific regularities, logic, strategy, self-evidence and 'reason.'* It is a question of analyzing a

'regime of practices'"Foucault (1991, :75). A regime of practices work as an ensemble, with a logic internal to the practices themselves. They are in this sense not given by ideology or derived from the state or interest groups. This regime of practices emerge out of the support practices find in each other, and the result is what Foucault calls a 'dispositif' (apparatus), a structure of flexible and contingent but relative stable relationships between practices. It is this structure that Foucault wants to decipher, taking account both of the contingency of these relationships, and their relative stability and durability. It is, however, crucial to understanding, that far from being located at different 'levels,' the one before the other, action and discourse are inextricable. Foucault is interested in what discourse does, not simply in what it says: how it forms the objects of which it speaks. It is not only because action on a particular object requires knowledge of that object, but also because discourse is itself practical, and not simply speculative.

Even today the logic behind prisons is not a unified, coherent whole, but rather an assemblage of distinctive and simultaneously intertwined logics which exists at the same time. This thesis claims that just like the prison is still surrounded by several coexisting interpretations As we can see today, there are several acceptable reasons for a society to make use of imprisonment. One focus is to cause general deterrence, aiming at reducing the likelihood that people will engage in criminal activity. A second approach focuses on restoring justice to victims by punishing those the offender. A third approach views the prison as a correctional facility constructed to impose personal reform upon those who are sent to prison. They are to be rehabilitated to function in society after release. Finally, it is a simple necessity to protect the public from those who have committed crimes or detain those who persistently follow a criminal path. All or some of these justifications may apply to the functioning of prisons. It seems as if there is an increased focus on reforming criminals to the extent that some prisons are called correctional institutions(Goyle, 2017). In similarity, nudging will be investigated through this line of thought. As can be seen with the prison nudging is still surrounded by several coexisting interpretations.

The Influence of Nudging Viewed in Light of Foucault

Foucault had an exceptional understanding of what characterize power-relations in a modern liberal society. Power for Foucault is neither repressive nor prohibitive. Power is not even opposed to freedom, on the contrary, as expressed by Foucault himself:

"It seems to me that power must be understood in the first instance as the multiplicity of force relations immanent in the sphere in which they operate and which constitute their own organisation; as the process which, through ceaseless struggles and confrontations, transforms, strengthens, or reverses them; as the support which these force relations find

in each other, thus forming a chain or system, or on the disjunctions and contradictions which isolate them from one another (...). Power is everywhere, not because it embraces everything, but because it comes from everywhere” (Foucault, 1990, pp.92-93).

According to Foucault, power can only be exercised if individuals can act autonomously and have the capacity to do so. The contemporary liberal state’s exercise of authority is a productive influence; it seeks to stimulate and develop skills in its citizens for the mutual benefit. The exercise of power is productive, in the sense that it works through individuals making them capable of transforming themselves into a specific subject (Mik-Meyer & Villadsen, 2007). In similarity, (Sunstein & Thaler, 2008) argue that *“Nudges are everywhere, even if we do not see them. Choice architecture, both good and bad, is pervasive and unavoidable, and it greatly affects our decisions”* (Sunstein & Thaler, 2008, p.253). In short, Thaler and Sunstein suggest, similar to Foucault that unrestricted autonomy in a pure form is a meaningless concept since *“choice architecture and its effects cannot be avoided”* (Sunstein & Thaler, 2008, p.79). In this regard, Thaler and Sunstein seem to agree with the Foucauldian conception that power is everywhere. Foucault departs in degree from Thaler and Sunstein by emphasizing that no one, in fact, holds power. Power is a somewhat independent force separated from the person in a position of power. Similarly, individuals who are in a position to influence, or who nudges other people, might be unaware of the power they hold either through intended or unintended actions. As Thaler and Sunstein state: *“Many real people turn out to be choice architects, most without realizing it” Sunstein and Thaler (2008, p.3).*

For Thaler and Sunstein, nudges are entrenched within the context of choice architecture. Similarly, Foucault discusses instruments of the state operating as an intervention, that does not openly prohibit certain unwanted behaviors, but strongly discourages it via the use of “adjacent, detective, medical and psychological techniques” while encouraging beneficial behavior (Foucault, 2007). Foucault analysis of the English philosopher Jeremy Bentham’s (1748-1832) ‘Panopticon’ is highly relevant for this thesis since it is a very telling example of a Foucauldian examination. Furthermore, the criticism of panopticon and nudging have significant similarities, both accused of making use of techniques that disregards the autonomy of those influenced. The panopticon is essentially a construction originally intended for prisons and other institutions which allowed for the perfect exercise of discipline.

“A new mode of obtaining power of mind over mind, in a quantity hitherto without example: and that, to a degree equally without example, secured by whoever chooses to have it so, against abuse” (Bentham, 1791, p.3).

The concept of the Panopticon building described a design which allowed a single guard to observe inmates of the prison without them being able to tell if they are being watched or not.

It would be impossible for one single guard to look at every single individual at once. The genius of this type of building stays in the fact that the inmates do not know for sure if they are being watched or not, hence they will act as if they are watched all the time, controlling their behavior constantly. An important aspect of nudging is the importance of obtaining a delicate understanding of what counts as an influence as opposed to coercion. The ability for resistance is fundamental, both within the Foucauldian logic of power and if nudges are to function properly, they have to be 'easy' to resist.

Dissecting Nudging Theory Into its Constituting Parts

Through the application of Foucault's ideas, this thesis will analyze nudging by splitting up the various parts that make up nudging as a unified whole. Nudging, however, is not an empirical object as such, but rather a method to influence people. The empirical object in this thesis is the literature/theory behind nudging. By applying Foucault, this thesis emphasizes the importance of distinguishing the individual parts making up the practice of nudging to identify its many theoretical parts and enables the researcher to point out the unavoidable intertwined logics upon which the practice of nudging rests. Foucault's theorizing on the constitution of the subject is integral to the critical and constructive reading of the field. As the thesis progressed back and forth through the emerging themes and patterns in the literature, it became increasingly natural to describe the progress in characterizing nudging, as a constant interplay between economics and psychology. Foucault's conceptual tools allowed reading of the articles enabled the identification of several factors contributing to curious inconsistencies in the construction of a coherent nudge framework.

Exposition of the Thesis and Choices

The message of this thesis is both critical of nudging and constructive. It is constructive in the sense that it points to conditions upon which we can accept nudges, these can be very hard to meet in practice. It is critical in that it develops the notions of nudge which makes explicit the tasks for both theorists and practitioners invested in nudging. A short description accompanies each chapter. The organization of the thesis begins with a short introduction to what the main goal is for each chapter.

Chapter 1 reviews the most important scholarly literature from the on nudging of the last ten years. This chapter begins with a brief introduction of behavioral economics, libertarian paternalism, and nudging. The next part of this chapter will present two common lines of critique inherent in the nudging literature as well as a brief reply by Thaler and Sunstein. The views are presented to give a brief overview of nudging to make the following chapters more relevant.

Part 1 proposes, from the discussion in chapter 1, that the misconception in the scholarly

literature are so severe that they intercept the conditions for general progress. This part will provide a concise explication of the foundational assumptions inherent in the current application of nudging. The two chapters in this part describes the two theoretical legs upon which nudges are composed. Chapter 2 introduces the scientific foundation behind nudging, and how decision-making is understood in dual-process theory and compliment with examples. First, this strategy ensures that the arguments developed in this thesis are relevant to the nudging literature. Most notably this chapter enables the thesis to raise questions about the ethical implications nudges. Chapter 3 begins by crystallizing how neoclassical economics and behavioral economics differ in their understanding of human agency. To make the difference clear three approaches to behavioral change is presented. Outlining these three methods sets the stage to understand the novelty of nudging, and why nudging it can be an attractive tool in public policy. Finally, this chapter presents the potential tension between nudging and essential aspects of conventional welfare economics. This chapter provides clarity, as to why Thaler and Sunstein purport that nudging is freedom preserving. If nudges are denied this quality, then their novelty and appeal to public policy would dramatically decline. The thesis continues with the comprehensive discussion of the various aspects and issue which must be taken into consideration before one can apply nudging.

Part 2 is picking up the current discussion in light of part 1's exposition of nudges, to get an in-depth analysis of the implications inherent in nudges in line with its proponent's views. Chapter 4 brings the insights from Thaler and Sunstein's book *Nudge - Improving Health Wealth and Happiness*, together with the prevailing assumptions and opinions within the discourse on nudging. This Chapter comments on the definition of nudging and explains why the definition is important in relation to its applicability, by bringing into play the information provided in the previous sections. Toward the end this chapter approaches a more critical take on nudging and sets the stage to raise questions about the ethical implications of nudging. Chapter 5 goes head to head with the implication of autonomy and nudging, Thaler and Sunstein claim that nudging is everywhere, but they do not go into depth about the difference in intentionally designing an environment, compared to a situation that just happens to be in place. This crucial difference is analyzed in this part. For numerous reasons, the terms of rationality, autonomy, and liberty are of great importance for the analysis of the debate on nudging. First, it has to be investigated to what extent the behavior of the individual citizen is influenced and whether there is any room left for notions of autonomy and liberty. It is somewhat questionable whether the nudged individual still maintains his/her capacity to act, as is greatly depends on the concrete nudge.

Chapter 6 analyses the ethical implications relevant to the application of nudging to public policy. Furthermore several other implications to evaluating nudges are analysed in the chapter, such as politics and transparency its current stage and questions to behavioral findings.

1 Chapter. Exposition of Propositions in Nudging and Critique Thereof

This chapter presents the most general preceptions of nudging in the literature and serves the purpose of identifying the most critical disagreements in the literature to provide a base for the later solutions developed in this paper.

1.1 Political Ideal of Libertarian Paternalism

Thaler and Sunstein's book on nudging is intertwined with their earlier work *Libertarian paternalism is not an Oxymoron* C. R. Sunstein and Thaler (2003). In this article, Thaler and Sunstein argue for a policy approach informed by behavioral insights C. R. Sunstein and Thaler (2003, p.4) Libertarian paternalism is placed in the middle ground between libertarianism and paternalism. In general, Paternalistic policies try to protect individuals from themselves. More precisely a paternalistic policy is an "*Intervention in a person's liberty of action which should serve that person's good, but takes place against or without her will*" (Fateh-Moghadam & Gutmann, 2014, :384). In contrast, libertarianism can be described as "*a political philosophy that affirms the rights of individuals to liberty, to acquire, keep, and exchange their holdings, and considers the protection of individual rights the primary role for the state*" (Vallentyne & van der Vossen, 2014, p.1). Another important trait of libertarianism to be highlighted is a focus on negative liberty as the "*absence of forcible interference from other agents when one attempts to do things*" Vallentyne and van der Vossen (2014, p.1). Combining these two political thoughts, Thaler and Sunstein propose that policymakers should consider how to mitigate human cognitive biases by making use of 'choice architecture,' as a third way in policymaking informed by behavioral economics. Libertarian paternalism describes the process of guiding individuals in their decisions while maintaining or even increasing their freedom of choice (C. R. Sunstein & Thaler, 2003, p.4). The Thaler and Sunstein state: "*Libertarian Paternalism is a relatively weak, soft, and nonintrusive type of paternalism because choices are not blocked, fenced off, or significantly burdened*" (C. R. Sunstein & Thaler, 2003, p.4). Nudging is presented as a tool that both posit a significant influence without explicitly preventing any options for the individual. Furthermore, if the recipient of a nudge behaves in accordance with the intended choice architecture, then one could simultaneously assume consent from the nudged. After all, nudging does not restrict our freedom of choice. This presentation might be a little crude, and it is the delicate nuances to this sentence that this thesis is going to explore.

The central point in Thaler and Sunstein's argument is the importance of guiding people to act in their self-interest and "*make choosers better off, as judged by themselves*" (Sunstein & Thaler, 2008, p.5). Less than a decade after the publication of (C. R. Sunstein & Thaler, 2003)'s

book, libertarian paternalism have become increasingly popular in political circles, and the nudge approach has gained major recognition, particularly in the western liberal democracies of the United States and the United Kingdom. In the United States, behavioral findings have provided an important reference point and informed many initiatives in various areas, involving fuel economy, energy efficiency, and healthcare (C. R. Sunstein, 2014). (Sunstein & Thaler, 2008)claims that nudging has four advantages compared to more conventional policies, such as bans:

1. People’s situations are highly diverse. Nudging reduces the costs of one-size-fits-all solutions by allowing people to decide for themselves.
2. Public officials will make mistakes. Errors with nudges cause less damage than bans because people remain free to ignore them.
3. Public officials may act under the influence of private groups. If their interests are met, it is a safeguard that people can go their way.
4. Restricting the freedom of individuals causes frustration and anger. When a government applies nudging it treats freedom as an intrinsic good and influence people through means of respect because individuals are left free to choose.

1.2 System 1 and System 2

Nudging theory is based on two important insights about human behavior and cognition derived from the work of (Kahneman, 2003)’s distinction between humans and homo economicus. Kahneman describes how human’s reflective and automatic systems impact human decision making (Kahneman, 2003, 2011). These works portray how humans act in irrational ways by misjudging or ignoring probabilities and misinterpret information. Furthermore, they show that these irrational tendencies can be quite predictable. This line of thinking can be considered a research program that studies persistent deviations from standard expected utility theory, by drawing on insights from psychology. For example, it has been found that: real-world decision-makers simplify choices they are represented with by using heuristics instead of calculating possible benefits (Kahneman, 2003). These studies are focused on providing lessons for individuals to accept and learn from their irrational tendencies. According to Kahneman, there is a need to distinguish between two kinds of thinking:

“System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control” (Kahneman, 2011, p.20)

“System 2 allocates attention to the effortful mental activities that demand it, including complex computations. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration” (Kahneman, 2011, p.21).

Thaler and Sunstein refer to these modes of thinking as respectively automatic thinking and reflective thinking (Sunstein & Thaler, 2008, p.19). The question of approval or disapproval of nudging depends on the specific preconception of autonomy and choice held by the author to determine how nudges exercise influence over individuals. To properly engage with this debate, this thesis will present and analyze the theoretical requirements for a successful nudge.

1.3 The Libertarian Critique

Nudging is interpreted numerous times in the literature as too paternalistic and restrictive regarding individual liberty. In line with this view, the first criticism focuses on dual process theory, the concept of bounded rationality and a core assumption in Thaler and Sunstein nudge. (Hausman & Welch, 2010) find it wrong to understand human decision making as systematically wrong, rather than simply 'occasionally' wrong.

Furthermore, this line of critique is skeptical whether it is meaningful to aim at guiding human behavior according to their true preferences. The motivation for this critique lies in their understanding of 'pure preferences' to be identical with the preferences of homo economics. By understanding 'pure preferences' as basically the same as rational choice, Hausman and Welch finds the addition of 'pure preferences' redundant. The implication is that there is no sure way of knowing what is deemed rational, by the agent's themselves, within every contingent situation. Attempts by policymakers to assess decisions of such an entity do inevitably lead to policies: *“Substituting the policy maker's judgment of what is good for that of the agent”* (Hausman & Welch, 2010, p.129). With this claim one does arrive at the formerly defined paternalism, being an *“intervention in a person's liberty of action which should serve that person's good, but takes place against or without her will”* (Fateh-Moghadam & Gutmann, 2014, p.384). Similar to this critique (Sugden, 2009) states that it is meaningless to refer to true preferences since all these do not have an objective, measurable definition (Sugden, 2009, p.370). It is, in general not possible to perceive such true interests, even if those would exist, claiming that Thaler and Sunstein do *“provide very little guidance about how she [the choice architect] is to discover those judgments”* (Sugden, 2009, p.367). It is a good critique, since identifying the preferences of the recipient is necessary if one claims to help them improve their choices. (Sugden, 2009), goes further and claims that even in the absence of bounded rationality, there is no reason to

believe that individuals can produce coherent preferences: that is, according to Sugden, we cannot assume that there is a rational economic agent 'inside' each human.

Thaler and Sunstein are also accused of failing to show "*that those people are making bad choices as judged by themselves*" (Sugden, 2009, p.371). Oliver is also skeptical about 'pure preferences.' He suggests, instead of speculating about possible underlying preference structures, policy-makers should simply accept performed choices as expressions of individuals' interests, as "*the best approximation of what a person really wants, (. . .) may in fact be given by what they choose*" (Oliver, 2013, p.11). If an individual, for example, does value the freedom to make mistakes, a nudge dissuading him/her from this behavior will express a general preference of rationality over the freedom of choice Amir and Lobel (2008, p.2120). Supporters of Thaler and Sunstein's nudging concept are accused of promoting rationality as a worthy, if not the highest cause itself, without considering that the preferences of nudged persons may deviate from that view (Amir & Lobel, 2008, p.2120). Concerning the topic of liberty, libertarian critique does show a variety of differing views, ranging from the claim that nudges necessarily do infringe upon liberty to the statement that they may, but do not have to violate liberty. It is stated that, for example, the setting of defaults itself would convey a certain normative message (Amir & Lobel, 2008, p.2121).

Nudges such as 'Save More Tomorrow' (Sunstein & Thaler, 2008, p.103) may be accused of encouraging a clear and subjective opinion regarding the question of how money should be invested. Such an influence is not referred to as open coercion but defined as a technique that "*does not interfere with a person's options*" but instead "*perverts the way that person reaches decisions, forms preferences or adopts goals*" (Wilkinson, 2013, p.344). This critique might be too harsh since far from every exogenous influence on the decision-making process can be regarded as manipulation. In general, to qualify as manipulation, an action has to involve an intentional actor causing or encouraging a decision, by withholding information leading to a process within which a rational person would in retrospect want to chose otherwise (Wilkinson, 2013, p.345). Wilkinson and several other authors in the group of libertarian critics use the negatively connoted term of manipulation, an influence that "*subverts and insults a person's autonomous decision-making*" (Wilkinson, 2013, p.345), expressing a strong rejection of nudge interventions. Despite their harsh position, they do acknowledge that nudges are yet not necessarily manipulative. If there is a genuine desire for improvement by the 'choice architect,' that is, "*if the nudger sincerely wants the targets not to act in the nudged way if the nudging is unsuitable for them*" (Wilkinson, 2013, p.354), then the nudge is not manipulative. The argument of suppression biases is viewed skeptically, as it is suspected that this suppression might have unintended effects that may result in "*a myriad of new predictable and unpredictable rational and irrational behaviors*" (Amir & Lobel, 2008, p.2116). It has to be added that some authors such as Hausmann and Welch are aware of the pejorative connotation and therefore use the term

'shaping' analogously to nudging. The term 'autonomy', frequently used as a more far-reaching category than the mere freedom of choice, is defined by Hausman and Welch as "*the control an individual has over his or her own evaluations and choices*" (Hausman & Welch, 2010, p.128). The term of autonomy differs within this context from the notion of liberty, as it is not only concerned with the preservation of a set of choices, but with the nature of the decision-making process itself. If behavior change is induced by any means but rational persuasion, for example by subliminally conveying information then, "*their autonomy is diminished*" (Hausman & Welch, 2010, p.128).

With a terminology close to Foucault, Bovens asserts that "*when we are subject to the mechanisms that are studied in the science of choice, then we are not fully in control of our actions*" (Bovens, 2009, p.4). Bovens is very critical towards the influence of nudges. When someone is nudged he understands it as being subject to a foreign influence or power, which may bear a possible long-term effect of infantilization (Bovens, 2009, p.11). Within an environment that increasingly rules out the possibility of making mistakes, it is no longer necessary, either engage in deliberation or to train our capacity for judgment. The development of individual's character is, by Bovens, considered endangered. He asserts that non-autonomous preference changes via nudges, induced by choice architects, may lead to an incoherent preference structure. Bovens coins this phenomenon as the creation of persons with a "*fragmented self*", "*incapable of taking their lives in their own hands*" (Bovens, 2009, p.14).

1.4 The Anti-Libertarian Critique

This line of reasoning often acknowledges the empowerment aspect as proposed by Thaler and Sunstein their worry is rather to confine nudging to areas where this empowerment is evident. These critics partly acknowledge Thaler and Sunstein's assumptions regarding the nature of man, "*humans regularly make systematic and predictable errors of judgment*" (Smith & McPherson, 2009, p.324). Contrary to the libertarian authors in the section depicted afore, Smith and McPherson take a different stance and claim that certain nudges may promote individual liberty. Dismissing a merely negative definition of liberty, they differentiate between formal liberty, being the "*absence of formal constraints on an individual's options,*" and substantive liberty, describing the "*opportunity for autonomous reflection*" (Smith & McPherson, 2009, p.330). Their proposition of substantial liberty is a more positive and far-reaching definition of liberty. Individuals have substantial liberty "*if they have the time (. . .) to reflect on their goals and aspirations and engage in practical reasoning and action toward the same*" (Smith & McPherson, 2009, p.330). For a nudge to be promoting substantive liberty, it has to effectively enhance people's "*opportunity or capability of autonomous reflection*" (Smith & McPherson, 2009, p.331).

Thus, nudges have to empower individuals, by encouraging autonomous deliberation. If a

nudge however, only focuses on increasing an individual's welfare without such an empowering element, it is evaluated as a paternalistic restriction of liberty. (Saghai, 2013) similarly agrees with the claim that nudges may preserve liberty, yet he decides to use the more narrow term "*freedom of choice*" (Saghai, 2013, p.488). To do so, nudges do not only need to preserve the pre-existent set of options but also have to fulfill the "*condition of substantial non-control*" (Saghai, 2013, p.489). This condition is met if the nudge is effortlessly or at least easily resistible, thus being a technique that counteracts biases but still does not include coercion. If this resistibility is unavailable, the nudge is considered to be controlling and thus restricting individual freedom of choice. In other words, if this condition of substantial non-control were not given, a nudge would exhibit full control or direct coercion of the nudged.

Included within the literature, is the British Professor of Political Science and Public Policy Peter John with a highly optimistic view regarding the nature of man and an alternative concept to nudge, namely 'Think' (John et al., 2011). They act on the assumption of knowledge-hungry individuals, eager to learn to process more information and attempting to enhance their capacity for reflection and judgment (John et al., 2011, p.19). Given enough time, information and an appropriate environment, citizens may come to the optimal judgment for themselves and others (John et al., 2011, p.11). Hence, empowering individuals to enable them to make good decisions by themselves is the main goal of the 'Think' strategy. For deliberative critics, the process of decision-making has a unique impact on the final choice. Regarding the topic of rationality, they state that instead of focusing on preexistent preferences, the way in which processes shape preferences should be at the center of attention (John et al., 2011, p.12). Instead of aiming to foreign influences on the person making a decision, the 'Think' strategy proposes an alternative to nudging with a technique that promotes discussion among citizens.

They hope that "*democratic deliberation can lessen the problem of bounded rationality*" (John et al., 2011, p.13). Autonomy is not interpreted as the absence of foreign influences but as an 'educational effect' initiated via contact and deliberation with other individuals (John et al., 2011, p.364). Contrary to the assumption of fixed preferences the deliberative critics illustrate a "*transformation of (often ill-informed) preferences via mutually supported deliberation*" (John et al. (2011, p.364). The Think approach, similar to nudging, perceives people as suffering from bounded rationality. However, instead of trying to "save" individuals the effort of thinking through their options. This method stimulates the individual's tendencies for critical thinking.

1.5 Common Reply by Thaler and Sunstein

(C. R. Sunstein & Thaler, 2003, p.175) have responded to these critiques by arguing that anti-paternalistic views are based on a false assumption and at least two misconceptions. The first misconception is that paternalism must be coercive, while the second is that paternalism has valid alternatives, and the false assumption is that people make choices that are better than the selection that would have been made for them by paternalism. Advocates of liberal paternalism

respond that nudging keeps all the options available to the individual. That paternalistic expert have a better understanding than individuals and can, therefore, promote better choices. Furthermore, since policy-makers must make a choice that will affect the behavior of others, it is best to do so in a way that aims at benefitting individuals, rather than staying ignorant of behavioral insights. Knowing that a neutral choice architecture design does not exist, it seems better to use current knowledge to improve choice designs rather than simply ignoring its influence on choice and hope for the best. This literature review was constructed to give a concise overview of some of the general debates surrounding nudges. *“By claiming that so many things are nudges, the approach, it could be argued, is in danger of becoming theoretically empty”*(Oliver, 2013, p.10) Based on the various views in chapter 1, it seems as if Thaler and Sunstein’s concept, nudge have lead to confusion and could benefit from increased clarity.

PART I. Exposition of Theoretical Propositions in Nudging and its Reference to Welfare Economics

Part 1 is made up of two chapters. Chapter 2 analyses dual process theory and how bounded rationality is understood. Chapter 2 discussion the implications of measuring welfare with the model of man developed in dual process theory and its implications to nudging.

2 Chapter. Exposition of Dual process Theory and Decision-Making

2.1 Starting Assumptions Bounded Rationality

Dual process theory claims that individual decision-making can, to a great extent be understood as a duality between two varieties of thinking: one, which is intuitive and automatic, and another, which is reflective and rational. (Kahneman, 2011) follows the proposition from the psychologists Keith E. Stanovich and Richard West who labels these two types of thinking ‘System 1’ and ‘System 2’ respectively. In dual-process ‘System 1’ is characterized as a fast, automatic, processing system that heavily relies on associations learned through experience. This ‘System 1’ is contrasted with ‘System 2’ a slow, deliberative, conscious processing restricted by a working memory system of limited capacity (Evans, 2008).

Cognitive models states that the interpretation of a situation often, expressed in automatic thoughts, influence people’s subsequent emotional, behavioral, and physiological response (J. S. Beck, 2011). While the results of behavioral economics and related fields are diverse, some ideas are shared across a wide range of subfields. Based on this common agreement, decision-making is described in a way that attributes different elements of a decision process to either one of the two systems. Dual process theory is to be considered a framework to ease the understanding of psychological findings. It is not a single theory in psychological decision theory, rather it is a general framework (Bargh & Chartrand, 1999), introduced to capture the core structure of research in psychological decision theory, bounded rationality, neuroeconomic, and behavioral economics broadly construed Bargh and Chartrand (1999). The dual process frameworks describe the automatic system as fast, effortless and emotional, thus best equipped to deal with simple guidelines and stimulation. “*Automatic thoughts are a stream of thinking that coexists with a more manifest stream of thought*” (A. T. Beck, 1964, p. 14 in Beck 2011). Dual process theory asserts that individuals immediate interpretation of a situation is expressed in automatic thoughts. We barely notice these thoughts, but express them through emotions behavior and physiological responses. The automatic thoughts occur quite briefly, and we are scarcely aware of our interpretations of them, rather we experience emotions in relation to these

automatic thoughts. Individuals can, with just a little training easily bring these thoughts into consciousness and learn from them. (J. S. Beck, 2011, pp.75-76). The reflective system is described as controlled, effortful and neutral, thus better able to deal with concepts. The findings in psychological decision theory state that biases, heuristics effects are directly derived from the idea that our automatic system simplifies and often misrepresents information, thus leading decision-makers to hasty decisions which are not always the self-serving ones.

Contemporary psychology now acknowledges that individual actions are affected jointly by processes set into motion directly by one's environment and by processes initiated by acts of conscious choice (Chaiken & Trope, 1999). Thus, mainstream psychology acknowledges that we as human beings conduct ourselves by focusing attention towards conscious thought processes, and less controllable environmentally triggered processes. People are thus said to be influenced simultaneously by conscious (control) and nonconscious (automatic) processes (Bargh & Chartrand, 1999).

Research in behavioral science is now focused on mapping out when one process (control or automatic) takes precedence over the other (Bargh & Chartrand, 1999). The British health psychologist and professor Theresa Mary Marteau states *"There is an increasing recognition that both these systems are very important in explaining our behavior. Often they work synergistically, so they work together well. Sometimes they work antagonistically. This is one of the reasons why, while many of us have very good intentions, we often find ourselves behaving in ways that go against our intention."* (Theresa Marteau in Science et al., 2011, p.17). The insights from this line of thought have been used in the Dutch airport in Schiphol (Sunstein & Thaler, 2008, p.4). Here they have painted a fly in the male urinals. In this example, the fly becomes salient when we enter the restroom. The fly is captured by our visual processes performed by automatic thinking. When this happens, we become surprised. It is an unusual situation which attracts reflective attention. A direct consequence of this encounter either results in a decision to aim for the fly or not, either way, the presence of the fly enhances the likelihood of the person to concentrate on the current task of urinating. This nudge can be said to function as a mental reminder, by making the current task of urinating salient.

It is crucial to notice that the 'System 1' and 'System 2' thinking does not imply that a given behavior is maintained or occurs exclusively within one of the two modes of thinking. These two systems are integrated, and the concepts are directly devised to ease the explanation of how they function and interact. Automatic thinking operates somewhat autonomously, whereas reflective thinking activates with doubt and in a context provided by automatic processes (Kahneman, 2011). Automatic and reflective thinking often interacts, but the reflective system seems to depend on the automatic system, while the opposite is not necessarily true (Marchiori, Adriaanse, & De Ridder, 2017).

When 'System 2' is not overseeing 'System 1', people are most likely to make mistakes.

Currently, dual process theories are the foundation behind nudging. For example, decisions are often based on perceived popularity, attached emotions or familiarity, rather than a rational cost–benefit analysis. The main difference between nudging and other interventions such as education campaigns is that nudging is supposed to ease the choices we face every day without requiring too much effort because it avoids engendering behavioral change through conscious thinking. In the majority of our decisions, we do not have the time, motivation, or resources to think our decision through (Gigerenzer & Gaissmaier, 2011; Hofmann, Friese, & Strack, 2009; Kahneman, 2003, 2011). As a result, most behaviors are arising out of a hermeneutic web of learned habits, heuristic processes, associations, and ingrained responses, rather than logical and rational processes (Gigerenzer & Gaissmaier, 2011; Hofmann et al., 2009; Kahneman, 2011; Rumelt, 2012). Contrary to homo economicus, humans are considered: imperfect information processors, emotional beings who get easily influenced by the context, short-sighted, myopic, inconsistent, and cognitively lazy C. R. Sunstein (2014). The premise for this view is an understanding of behavior with concentration as a depletable resource which we spend throughout the day (Casey et al., 2011).

2.2 Human and Econs – Thaler and Sunstein’s Application of Dual-Process Theory

Drawing on dual process theory Thaler and Sunstein explain their idea of man, as a person with an internal division between ‘Econ’ and ‘Human.’ An individual may qualify as an ‘Econ,’ if they consistently behave ‘perfectly rational’ and are not ‘systematically wrong’ regarding their decisions (Sunstein & Thaler, 2008, p.7). Humans, in contrast, are defined by their predisposition to make ‘biased’ forecasts and decisions. They are bad at calculating probabilities; they exhibit biases by overweighting small probabilities and underweighting large ones, they are more risk-averse when considering possible gains than when considering potential losses (Kahneman & Tversky, 1979). Their choices are subject to the ‘endowment effect,’ (Daniel Kahneman, 2000) they commit ‘the conjunction fallacy,’ and so on (Kahneman, 2003; Starmer, 2000; Thaler, 1999). Thaler & Sunstein’s distinction between ‘Humans’ and ‘Econs’ focuses on the cognitive and rational capacities of individuals. While ‘Econ’s only rely on their reflective system, ‘Humans’, in general, act exclusively through their automatic system, because relying exclusively on the reflective system is infeasible (Kahneman, 2011). Reflecting can be time-consuming and inappropriate in most of our daily activities (Kahneman, 2011). Therefore, individuals will in most cases act according to Thaler and Sunstein’s archetype ‘Human’. Although the intuition of the automatic system swiftly promises seemingly satisficing decisions, it is equally prone to systematic errors, such as ‘Anchoring,’ ‘Availability,’ ‘Representativeness,’ ‘Overconfidence,’ ‘Loss Aversion,’ ‘Status Quo Bias,’ and ‘Framing’ (Sunstein & Thaler, 2008,

pp.23-37).

The study of bounded rationality has established several firmly held beliefs about aspects that influence human behavior, some of the most common will be presented here. 'Anchoring' is the tendency to rely heavily on the first piece of information encountered in a decision context. Hereafter other judgments are carried out by adjusting to the first encountered information (the anchor). Humans tend to choose inappropriate anchors and to adopt a nearly random scale, resulting in poor estimations, in purely economic decisions. (Sunstein & Thaler, 2008, p.19)citing(Jones-Lee & Loomes, 2001, p.210). 'Availability' describes the phenomenon of humans assessing "*the likelihood of risks by asking how readily examples come to mind*" (Sunstein & Thaler, 2008, p.25), leading to a severely biased evaluation of risk and in consequence to irrational risk-related behavior. 'Representativeness' Irrational behavior can also be a result of human tendencies to identify patterns in random outcomes, based on these perceived 'trends' we proceed to make illogical conclusions (Sunstein & Thaler, 2008, p.27). Reports of airplane accidents and train derailments often lead people to believe that such events are much more typical than they truly are leading them to avoid certain actions for irrational reasons. Also, internal factors such as optimism and overconfidence produce these effects (Sunstein & Thaler, 2008, p.31). 'Loss aversion' refers to the tendency of individuals to dislike losses more than they appreciate gains, which more concretely means that there is a nonlinear relationship between the two when determining personal welfare (Kahneman & Tversky, 1979).

2.3 How to Identify Human Biases

The concept of 'script,' as elaborated by (Akrich & Latour, 1992), might help to explain the functioning of a nudge. A script is built-in prescriptions that impose themselves in the choice context, inviting one choice of action over another. Behavioral scripts are a sequence of expected behaviors for a given situation. Scripts include default standards for the actors, props, setting, and sequence of events that are expected to occur in a particular situation. Before we can implement a nudge in a restaurant, one would have to analyze the different components in the restaurant. We would look at the expected roles and functions of table decoration, menu, food, and, the customer, waiters, and chefs. Identifying the scenario enables one to create a sequence of planned events for this script. Most likely it begins with a hungry customer entering the restaurant, ordering, eating, paying and finally leaving. Individuals frequently follow scripts; these are acquired through habit, practices and simple routine. By identifying this scenario, one can help people to save time and mental effort of deciding on appropriate behavior familiar situations. A nudge can be applied after the behavioral script is identified, to enhance the likelihood that the script will go as planned, by increasing the intuitiveness of the menu. One can imagine that it would be tempting for a restaurant to design nudges which can enhance

the consumption by costumers (Spence, Harrar, & Piqueras-Fiszman, 2012). Identifying scripts brings us to the design of defaults.

2.4 The Influence Exerted by Default Choices

One of the most influential biases the authors present is the 'status quo bias,' which can easily be explained as humans' "*tendency to stick with their current situation*" (Sunstein & Thaler, 2008, pp.73-76). When structuring choices like the design of retirement plans, the installation of a default option has an astounding influence on the choices humans make (Sunstein & Thaler, 2008, p.83). Because to the fear of committing a mistake and due to a general inertia, humans tend to stick with the default instead of making an active choice (Ilana Ritov, 2000). In addition to recommending an option, the default determines a standard choice that applies in cases where the individual does not express a preference.

Installing a default helps individuals in situations with complex options. The nudge of giving feedback is intended to provide humans the opportunity to learn from their decisions (Sunstein & Thaler, 2008, p.90), by directly administering information about the individual's prior choices. The provision of information is the technique of visualizing options. According to (Sunstein & Thaler, 2008), people are often unable to link options to possible welfare outcomes. By providing an accurate depiction of each option's consequences may resolve individual's uncertainty regarding the option that will serve them best (Sunstein & Thaler, 2008, p.91). However, by allowing for such a broad characterization of nudges, Thaler and Sunstein start to blur the boundaries of nudging. Applying the insights from *Nudge - Improving Decisions About Health, Wealth, and Happiness*, we learn that status quo bias, inertia and the endowment effect all suggests that people tend to stick to the default options presented to them(Sunstein & Thaler, 2008, pp.7-8). According to this book, the default rule should aim at welfare maximizing (Sunstein & Thaler, 2008, pp.4-5). This objective illustrates a skew towards paternalism. Setting people as organ donors by default rely heavily on the status quo bias, as well as framing.

2.5 The Power of Framing

When individuals look at complex problems, they try to focus on the essentials and to screen out irrelevant distractions. However, the way we identify the most relevant factors and the best way to respond varies greatly with the framework adopted. People tend to treat frames as a given when people are presented with a frame; they get locked into a narrow perception and lose perspective. If the frame is inappropriate, then the quality of decision-making will often be reduced as well (Russo, Schoemaker, & Russo, 1989). People are prone to develop myopia

about both the options and the criteria under consideration. When a problem is framed in a particular way, people understand and analyze it with a limited set of plausible solutions; the frame is unwittingly closing off other possibilities.

A simple example is shown in figure two. Here 13 is perceived as B within a context of letters and B is seen as the number 13 when displayed within a context of numbers. 'Framing' tries to capture the effect of conveying information with a particular object in mind. An ad for a high-end automobile mentions words like "luxurious" and "comfortable," while avoiding terms such as "expensive" and "costly." A similar ad for a budget car would use words like "affordable" and "reliable" instead of "cheap" and "simple." Framing can have a significant influence on our behavior (Sunstein & Thaler, 2008). More generally, expectations are a powerful determinant of accessibility. More interestingly this figure also illustrates the complete suppression of ambiguity in conscious perception (see figure 1). As (Kahneman, 2003) states, "*This aspect of the demonstration is spoiled for the reader who sees the two versions in close proximity, but when the two lines are shown separately, observers will not spontaneously become aware of the alternative interpretation.*" (Kahneman, 2003, p.455) They see the interpretation of the object that is the most likely in its context, while the thought of it being any different is ignored. As found by (Klein, 1999), experienced decision-makers, are rarely affected by framing, within their field of expertise, in most cases', only a single option comes to their mind (Kahneman, 2003).

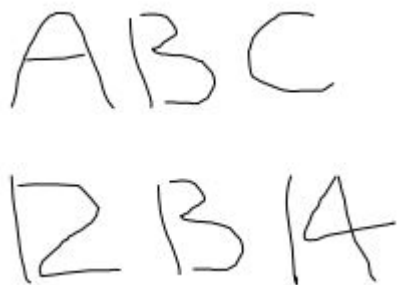


Figure 1: Figure 1 Ambiguity for perception

2.6 Social Biases

In contrast to these biases ingrained in individuals, social biases are rooted in the interaction of individuals within groups. 'Herding / Doing What Others Do' refers to the tendency of humans to defy evidence of their senses due to the influence of other individuals within a group, often based on a simple desire for conformity and avoiding the disapproval of the group (Sunstein & Thaler, 2008, p.56). 'The spotlight effect' happens when we think that 'everyone' is paying

attention to us. It is most effective when we are concerned with an attribute. For example, when an individual, experience a stain on their shirt, they assume that the attention of other people will be focused on the stain as well. In reality, it is unlikely that people will notice the stain, and far from the extent to which the individual himself experience it. The spotlight effect typically drives us towards conformist behavior describes the insight that humans, due to an excessive estimation of the degree of other people's attention, may tend towards conformist behavior if they have the feeling of being observed (Sunstein & Thaler, 2008, p.60). Two of the most common shortcuts are "*choose what everyone else has chosen*" or "*do the same thing I did last time.*" If behavioral evidence is ignored, the policy may not affect behavior the way it intended (Dolan et al., 2012). Moreover, the findings bear implications for information campaigns. Studies show that conscious thought has a minimal influence on much of our daily behavior. Instead, the behavior is often guided by 'automatic' responses to our environment and the things we encounter (Bargh & Chartrand, 1999).

The term 'satisficing' was coined more than 60 years ago by the American social scientist Herbert Alexander. Simon (1916-2001) describe, the tendency of people to select the first option that meets their minimum requirements rather than exercising maximizing behavior (Simon, 1955). Satisficing describes how individually and collectively, people have difficulty passing up the first reasonable solution. When a satisficing solution has been identified people often, unaware of the act, begin to analyze the pros and cons of their solution instead of seeking out other alternatives. The impulse to "*dive in*" makes sense for a decision that is acute, low stakes, and reversible. This approach is, however, inappropriate for reaching decisions such as signing a mortgage or buying a car. Salespeople often offer discounts which expire if we do not accept their offer right away. This technique relies on satisficing to make costumes make quick decisions. In the same vein, (Kahneman, 2011; Sunstein & Thaler, 2008) tells us that real-life decision-making is not as ideally and abstractly rational as in the theoretical accounts of rational choice, and several mechanisms, such as the ones mentioned above, have been found to persist across different choices. Instead of meticulously calculating the costs and benefits, we often use mental shortcuts or 'rules of thumb' to select an option that is 'satisficing,' although perhaps not the best (Kahneman, 2011). Behavioral findings have wide implications for public policy, by proving that some policies can be improved by incorporating the findings from behavioral economics. The ethical debate surrounding nudging has become more complex, and far from sufficiently explored. Nudging and dual process theory are accused of disregarding individuals capacity for decision-making.

This critique especially focuses on the fact that most of these experiments in dual process theory confine rational decision making within the sphere of logic and statistical models. However, as (Gigerenzer & Gaissmaier, 2011) points out, for many decisions, the issues are empirical rather than a priori issues. People do not expect that their test is designed to trick

them or show themselves as inconsistent decision-makers.

2.7 Habits and their Influence on Choice

Beyond that, a lot of the behaviors displayed in everyday life are acted out as part of regular routines of which we are only semi-conscious. Habitual reactions are likely to be accomplished when their triggering circumstances are encountered, due to the simple robustness of habitual representations in memory. The portrayal from habits is most intense and hence most likely to be activated compared to substitute responses. Furthermore, the alternative responses to the triggering circumstance may diminish as people persistently retrieve an identical response from memory, reinforcing its efficacy (Mc Culloch, Aarts, Fujita, & Bargh, 2008). Also contributing to habit performativity, people may misattribute externally-prompted representations with their natural response to the situation, that is, to their deliberate preferences and desires (Loersch & Payne, 2011). With this misattribution, habitual responses can seem to be constructed to reach deliberate goals and intentions. Even so, as the choice becomes habitual, the reaction becomes increasingly controlled by the triggering context other than the consequence of the habitual response. Thus, habitually reinforced habits can trigger people to peculiar actions, such as eating stale popcorn in the cinema, regardless of their taste and eat everything on their plates regardless of appetite. (Neal, Wood, Wu, & Kurlander, 2011; Wansink, Painter, & North, 2005).

The stimulus-cued nature of habits produces an increasing insensitivity to outcomes of the response.

“Habitual responses are initiated automatically in the sense that they are stimulus driven, require only minimal or sporadic conscious monitoring and thus are largely autonomous, and they do not depend on a goal for performance” (Wood, Labrecque, Lin, & Runger, 2014, p.5).

Despite the reinforcing characteristics of habitual responses. They remain largely controllable, especially if the individual is aware of their habit, or when brought to mind in a given context. The habit will often be resisted or changed with explicit goals incompatible with their habits. In these circumstances people will alter their responses in several ways (Quinn, Pascoe, Wood, & Neal, 2010). Habits may be initiated by a pursuit of a specific goal. After actions have been repeated over time, patterns in the choice context may trigger a habitual response intentionally or unintentionally. Habits might also develop without an attempt to reach any specific goals. The findings in (Kahneman, 2011) indicates that a lot of the things humans do on a daily basis

are performed on autopilot and may not always serve a conscious purpose. Habits are mediated by the automatic System 1 in dual process theory.

2.8 Reinforcing Self-Control Strategies

The operations of nudges may be understood to assist 'internal self-control strategies.' These are strategies or actions planned to ease possible situations in the future (C. R. Sunstein, 2000). Nudges are designed to both nullify adverse effects of cognitive biases and to serve as an external self-control strategy. Due to the ineptitude of the market to assist humans with their irrationality (Akerlof, 1991), Thaler and Sunstein expect choice architects to help them by promoting well-intentioned choice architecture within the public and private sector. Understanding why people act contrary to their beliefs is necessary to figure out what affects human behavior at large. *"To stumble is human. With every choice we make, individual motivation interacts with emotions, cognition, and social norms"*(Amir & Lobel, 2008, p.2098).

There are thus a plethora of aspects which affect our decision-making and whether one aspect takes precedence over another is very hard to predict. The American political scientist Sue Crawford describes social norms as shared understandings about actions that are either obligatory, permitted or forbidden (Crawford & Ostrom, 1995). One might not want to drink alcohol, but at a social event, fitting in might seem more important. Furthermore, social norms *"signal appropriate behavior or actions taken by the majority of people (although what is deemed 'appropriate' is itself subject to continual change)"* (Samson, 2014, p.27). Several factors are influencing individual behavior, and explaining why an action is performed can be challenging and predict it can be even harder. Nudges rely on human biases and try to exploit them in favor of the citizen's benefit. According to Thaler and Sunstein, nudges are in general necessary if conditions for a rational decision is restricted. If the benefits and costs of a choice are separated over time; are infrequent and complex and; if there is no feedback or opportunities for learning; and if the individuals do not have pre-established preferences, in all these cases nudging can and should be considered (Sunstein & Thaler, 2008).

2.9 Rationality

Cognitive biases incorporate consistent patterns of deviation from a baseline comparison . In this thesis, the baseline is interpreted to be made up by neoclassical assumptions of human behavior (Dutta, 1999). Since, behavioral economics tries to explain why we systematically deviate from the normative prescriptions of 'Homo Economicus,' by pointing out factors causing our behavior to be inconsistent by seemingly irrelevant factors (Simon, 1955). In general us-

age, the term rational is used to represent an ideal action as well as suggesting a normative (the correct) action, whereas irrational actions are to be prevented and can be considered morally dubious. Regarding behavior in the social sciences, rationality is taken to mean processes of thoughtful deliberation and reflection of different courses of actions and their likely outcomes.

In a technical, economic use, rational actions are synonymous with strictly private utility and self-interest. In the normative aspect, rational actions are often considered virtuous because they work against emotional (uninformed) responses to situations. Emotionally informed actions are on the other hand considered as an unpredictable endangerment of our autonomy leading to potentially harmful actions. By broad way of understanding rational decision-making is to a greater extent freed from the moral association and strict self-interest, implied in with strict substantive rationality. This Ecological rationality does not curb the individual from reflection, calculation and attempts to secure personal interests. However, it gives far more validity to emotional drivers for responses, such as empathy, without treating it as a biased decision. Using rationality in this sense has two advantages in the application of behavioral economics. First, it means that the rationale is not only a divisive self-interested one, which has to be monitored. Secondly, it brings to light our dependency to rely on emotional intelligence, and gut reactions to manage the negotiation of a variety of everyday situations. Looking at human behavior through these lenses significantly reduce our tendency to anoint ourselves, as more capable than individuals to select what would be an appropriate behavior for others.

3 Chapter 3. Welfare Economics – Combining Nudging and Preferences

This Chapter begins with presenting neoclassical thought. Presenting this logic is intended to show the clear division and similarities between the neoclassical and the assumption inherent in nudging since nudging is presented as a suggestive improvement of neoclassical models.

Economists distinguish between ‘positive economics’ and ‘normative economics.’ Positive economics is the study of cause and effect it attempts to create facts. It investigates questions such as “*If butter producers are paid a subsidy, will the price of butter be lowered?*” Normative economics, is political, it is concerned with public policy and trade-offs between desired and undesired effects: Should income be taxed progressively? When predictions occur, they are usually surrounded by a host of confounding extraneous factors or simple chance. Often economists have evaluated theories by their plausibility and emphasize intrinsic factors such as elegance and internal consistency. Value judgments, prejudices, and ideological biases unavoidably interfere with the conclusions espoused by the individual economist. The disparity between the confidence of economists place in accurate economic models, and the apparent disorderly natural world can be enormously frustrating to scientists in other fields.

3.1 Neoclassical Economics

The biggest distinction between behavioral economics and neoclassical economics is the one regarding the way humans choose (Mueller, 2004). Neoclassical economics sees the individual as the Homo Economicus, who are rational individuals capable of maximizing utility and act independently through full information (A. Sen, 1973).

The theory which informs neoclassical economics assumes that, for the time being, agents possess a static set of preferences over all likely outcomes. According to this theory, agents are rational if their preferences are complete, logically ordered without inconsistencies. This means that if a choice set is ordered as preferring $a > b > c$ then the agent will also prefer $a > c$. In uncertain outcomes, rational agents will prefer any option that gives them the highest net value, by computing the prospect of each event.

These individual preferences are expressed through mathematical axioms. A rational agent will follow these mathematical formulas in their pursuit of preference satisfaction. An individual’s choices are expected to maintain internal consistency between preferences. The rational choice theory holds that all considerations related to a choice, that is all possible considerations: emotions, norms or even religion, can be incorporated into the agent’s preference matrix. These preferences can be accessed indirectly by outsiders. By observing the agent’s actions, one infers from the observed behavior, the preference hierarchy of the individual.

Neoclassical economics attempts to explain and predict behavior by assuming rational agency.

Crudely stated, one could say that there are two views among practitioners. The first group mostly use the theory to represent or model individual behavior. The accuracy in describing individual's cognitive processes according to these principles of rational choice theory is somewhat irrelevant, as long as the model provides reasonable predictability (Freidman, 1994). These researchers maintain the model for its practical usefulness and are content with using the axioms of rational choice as long as it provides somewhat reliable results (Blaug, 2017). Their goal is not to explain behavior but rather to model actions and predict them. The value of the model is that when predicting aggregate behavior, it has a decent accuracy. This view model behavior, as if, people act according to rational choice, but they do not claim to have any insights into actual thought processes in the mind of individuals.

The second view maintains that rational actors do think according to the axioms of rational choice. This view upholds rational choice theory as a theory that describes the normative foundations of rational decision making. Many economists hold the view that rational choice theory is a powerful analytic tool precisely because it is internally consistent and reflects the actual principles which must characterize agency.

3.2 Neoclassical Economic Framework for Implementing Public Policy - Game Theory.

3.3 7.1 Homo Economicus and Humans

The biggest distinction between behavioral economics and neoclassical economics is the one regarding the way humans choose Mueller (2004). Neoclassical Economics sees the individual as the Homo Economicus, who are rational individuals capable of maximizing utility and act independently through full information A. Sen (1973). One of the most famous experiments from game theory will help to explain the neoclassical understanding of human behaviour. In short, there are two players in this game. The first one, 'the allocator,' is endowed with a sum of money, \$10 and has to offer some of the money > 0 to the second player, 'the recipient.' Now, the recipient has to make a decision, either accept or reject the offer from the allocator. If the recipient accepts tender, both players receive the money they won from the game. If the recipient rejects the offer, both players receive nothing. Based on the standard economic theory the recipient is supposed to accept any offer > 0 because the gain would always put the recipient in a better situation, compared to higher rejecting the offer. Moreover, it would be rational, for the allocator to provide the smallest possible amount to the recipient (\$1). In contrast with the theory, neither of the players behaves this way. Thus one of the basic assumptions of standard Economics, homo economicus, is violated by this very simple game. A common reply to these deviations is that the result should be seen as the analysis of a benchmark case, rather than a

description of actual game playing situations or a prescription for what rational players should do (Simon 2010 in Pacuit & Roy, 2017).

3.4 Behavioral Economics – Explaining Deviations from Neoclassical Economics

These deviations are often attributed to three different reasons:

1. According to Fong, Bowles, and Gintis (2006), the allocator has a fear of rejection, of his offer, which makes him more self-protective and less profit-seeking.
2. People prefer to act fair and resist inequality. As a consequence, they are willing to forego a gain to punish the allocator for assigning an unequal division of the reward. According to Fehr and Schmidt (1999), this phenomenon is called Inequity Aversion.
3. According to Francesco (2010), human beings are by default cooperative by their nature, as long as the other player acts in an altruistic way. It is first by experiencing lack of cooperation that people begin to act selfishly.

Behavioural economics criticise rational choice theory for maintaining unreasonable assumptions about the thought processes of human behavior. It argues that leaving out psychological processes to explain human behavior is insufficient. At the same time, behavioral economics endorses the normative view of rational choice. The function of psychology in behavioral economics is to give a psychological account of why people fail to behave according to the rational choice theory.

The presentation of behavioral economics does not cover all versions of behavioral economics in general. This thesis contends that the ideas behind nudging do incorporate this understanding of human behavior. By claiming to know the inner thought processes, what individuals 'truly' want, more than individuals themselves. Because of this claiming to know individuals' true preferences, one cannot be surprised that nudging interventions are suspected of manipulating people. If the framework behind nudging would claim simply to improve the prediction of human action by incorporating behavioral insights, then nudging is more likely to be considered acceptable. In behavioral economics, individuals can be assumed to act as if they were maximizing a subjective value function (Mueller, 2004). Experiments involving the ultimatum game have revealed significant differences across cultures in the sizes of the offers made by the first players and the willingness of the second players to accept these offers (Bergstrom, 2002). These experiments indicate the importance of social conditioning in determining individual behavior.

The salient feature of behavioral economics that differentiates it from mainstream neoclassical economics is that it relies on concepts and findings from psychology and other sciences to inform its assumptions about what it is that individuals maximize (Kahneman, 2003; Mueller, 2004). Behavioural economics will not abandon the premise of maximizing behavior. Their focus is rather to specify with greater precision what it is that is being maximized and why we sometimes fail to do so. I will come back to this point later in chapter 6, in the section preferences.

3.5 Three Approaches to Behavioral Change

3.5.1 1. Provision of Information - the 'Think' Approach

The 'Think' approach, aims to improve individuals capacity for decision-making by providing information relevant to the choice environment they are within and leave them free to choose what suits them best (John et al., 2011). It subscribes to the old democratic ideals to inspire public deliberation as an avenue to support free and equal citizenship (John et al., 2011). It is designed to avoid patronizing individuals and the accompanying implication that the government or civil society institutions prescribe what is best. It is discussable to what extent the information provided will consequentially lead to effective 'thinking.' That is how much actual behavioral change is brought about through information alone. John *“argues it is possible to get citizens to think through challenging issues in innovative ways that allow for evidence, and the opinions of all, to count”* (John et al., 2011, p.10). John argues that it is possible to get citizens to think through their problems and along the way, the broader challenges in society in a process that allows for the evidence, and opinions of all parties, to count. An example of this approach in the sustainability context is conservation education, for which the ultimate purpose is to influence individuals' behaviors and call for change in a way that supports conservation practices without imposing on people how they should act.

3.5.2 2. Direct Regulation - the Shove Approach

Direct regulation stands on the opposite side of the 'Think' approach in the continuum of behavior change strategies. It is often labeled as the 'shove' approach (French, 2011) and consists in using tools such as: economic incentives, mandates, and bans. One of the benefits of this strategy is that it brings fast results when it is put in place (Dobson, 2007). However, it can backfire in the longer run, because it makes use of coercion and restriction of liberty, based on an assumption of what is best for people or the general welfare. Furthermore, if out of line with public opinion there is a threat to endangering trust in the public sector, this approach can be seen as a last resort. *“Much of the time, and in many areas of policy, simply telling people*

what to do can be wholly counterproductive, especially at a time when the deterrence is low, and mistrust of politicians and civil servants is high." (John et al., 2011, p.4). It is also the case that governments in many modern industrial societies can no longer rely on deterrence and obedience to messages from a public institution, as they did before. Citizens will question the authority of the government or simply ignore it (Ilana Ritov, 2000; John et al., 2011). Nevertheless, some studies defend that when it comes to environmental issues, externalities might justify a mandate (C. R. Sunstein & Reisch, 2014), such as the case of putting a price on carbon.

3.5.3 3. Choice architecture - the Nudge Approach

In nudge theory, a choice architect is defined as anyone who organizes the environment in which people make decisions (Sunstein & Thaler, 2008). Thus, a choice architect is someone who indirectly influences the choices that other people make. Given that choice architecture cannot be avoided, the authors defend that the golden rule of nudging is to *"offer nudges that are most likely to help and least liable to inflict harm"* Sunstein and Thaler (2008, p.72). Choice architects need to understand how individuals behave and choose, to setup a choice architecture that will help people improve their capacity to select optimally. Thaler and Sunstein add: *"The potential for beneficial nudging also depends on the ability of the Nudgers to make good guesses about what is best for the Nudgees. In general, Nudgers will be able to make good guesses when they have much more expertise at their disposal"*(Sunstein & Thaler, 2008, p.247). Nudges use behavioral science to investigate and understand aspects of the human psyche and heuristics in decision making, such as biases and overconfidence (Kahneman, 2011). Drawing from the revealed 'failures' in human reasoning, the 'nudge' approach aims at setting up a choice architecture, which guides decision-making in a direction that should, ultimately, benefit the individual, even according to the individual's own beliefs. Even when individuals are willing to act more sustainably, people may not make choices and take actions that reflect this belief.

The separation between willingness and action may ultimately lead to economic and environmental harm (C. R. Sunstein & Reisch, 2014). The nudge approach has in recent years reviewed increasing interest, given its' applicability both in the public and private sectors. The nudge approach sets concrete objectives for the outcome of behavior change interventions and does not seek to influence values and attitudes. Libertarian paternalism sets up a direction for behavior while maintaining autonomy and choice for those directed. Thaler and Sunstein argue that the complexity of modern life and the speed of development and change *"undermine arguments for rigid mandates or dogmatic laissez-faire"* (Sunstein & Thaler, 2008, p.253). Thaler and Sunstein do not neglect the power of incentives and bans in changing people's behavior, but they do urge us to privilege the design of interventions that recognize citizens as boundedly rational decision-makers. They recommend that governments think carefully of default options

when they impose interventions on their citizen's choices.

3.6 Why Nudging is an Attractive Tool in Public Policy

Nudges should not change the structure or function of government programs or policies but alter the programs to operate at a lower cost or obtaining higher yield through behavioral insights. With this mandate, behavioral economics has become more than a theoretical exercise. Now dozens of behavioral nudges have been implemented on a massive scale, affecting the lives of hundreds of millions. A comprehensive list of such interventions would be difficult to compile and be distracting from the main points of this thesis. Many of these efforts have been summarized in the annual report of the (SBST, 2016). The success of these behavioral approaches has been paved through the promise of improving results without backlash. We can briefly state that by definition, a behavioral nudge cannot in an economic sense, alter the choice set but only modify the way in which the individual perceives the choice set. By refraining from altering the choice set, the decision maker would presumably have little reason to feel harmed by the nudge. If these requirements can be managed, it seems that true nudges will not create regret or other forms of infringements in its target. This section does not attempt to underplay the significance of recent behavioral insights. The goal is to warn against overestimating the ability of choice architecture to solve societal problems. One can question whether setting defaults, the primary libertarian paternalistic tool, is always the 'softest' tool in the policymakers toolbox, which includes such instruments as incentivizing, collaborative standard setting, education, and self-deliberation processes. Nudging is considered paternalistic because it improves individual welfare, and libertarian because opt-out has close to no costs.

3.7 Measuring Welfare Improvements from Nudging

Welfare economics is primarily about coordinating individuals, within an economy, or community. Individual minds and behavior are not the primary focus, except to the extent that improving individuals' choices can improve the system society, the economy as a whole. The primary purpose of welfare economics is to evaluate the overall impact of economic policy on the well-being of the affected individuals (British encyclopedia). We rely on assumptions to balance the trade-offs in welfare among various actors in the economy. These assumptions are used to derive some measure of overall well-being and determine which initiative satisfactorily can maximize welfare while imposing some other policy objective. Typically, several similar policies may be used to achieve the objective. In the case of biofuels; these could be taxes on traditional fuels; subsidies on biofuels; blending mandates, or caps on the production of traditional fuels. Each has the potential to achieve the objective of a minimum threshold of biofuel use. Depending on the approach the overall cost to consumers, producers, and taxpayers may be substantially (Just, 2017, p.3). While individual contexts can differ in important ways, there are

some general results from welfare theory that is important and have wide application. The two fundamental welfare theorems that relate competitive equilibrium to Pareto efficiency Hicks (1939); Kaldor (1939) are the most widely cited and (Just, 2017, p.4). If one is attempting to reduce CO2 emissions from manufacturing to generate greater welfare, then according to these welfare theorems the best option would be to tax emissions rather than taxing inputs or production during the manufacturing process. Taxing production, for example, can lead to production processes that produce more emissions but are cheaper for the relative quantities that prevail under the tax (Levin & Milgrom, 2004). While welfare theory considers all actors in a market or economy, this thesis will focus on the individual consumer. Much of behavioral economics is concerned primarily with the behavior of individuals, and, in the policy context, how context can influence the individual's behavior.

Consider the standard consumer problem: $\max_{x \in C} U(x | \Theta)$

where $U(\cdot)$ is the utility function representing the individual's well-being as a single dimensional valued function of their chosen consumption $\theta \in \Theta$. The context Θ represent all potential factors that could be altered by a behavioral nudge. The individual is restricted from selecting any choice outside $C \subset R_+^n$. This choice set is often just a representation of the budget available and the prices of potential goods Levin and Milgrom (2004). Regarding welfare, this model has simple implications for traditional policies. Traditional policies tend to change choice by changing the available choice set. If an economic policy makes a choice set $[C]$, larger, then the policy could potentially make the individual better off by supplying superior options. Instead of choice set $[A, B]$ a third option, $[C]$, has been added. According to economic theory, It is not possible that the individual welfare can be reduced by adding choices to the choice set $[C]$, because the optimal choice from the smaller choice set is still available, and presumably the individual would reject any choices that are inferior to this outcome. On the other hand, if the policy reduces the choices available in choice set $[C]$, then the individual is potentially worse off. A policy change that alters a person's behavior by reducing the choice set $[C]$ must make the individual, at least weakly worse off. Otherwise, the individuals affected by the policy intervention would not have changed behavior, as they were already selecting their optimal option in the choice set $[C]$.

This simple theory of the consumers underlies much of how welfare engages with policy on the consumer side. The key principle is that whenever a policy reduces the choice set $[C]$, there must be some benefit given to other actors in the economy to justify the restriction. In other words, whenever we suggest a policy restricting consumer choice, a burden of proof must be met to ensure that the policy has the potential to be welfare improving. This burden of proof is met by arguing that the individual's choices result in externalities affecting the well-being of the public who cannot directly influence this choice. When public planners are reducing externalities, they are reducing the choices available to one or several actors to enable a broader choice

set for the public Levin and Milgrom (2004). Sometimes, policies are primarily set up to save individuals from their own poor decisions, such as wearing a seatbelt. Many policies go along this type of policies with minor impact on others. Other policy areas in this realm of paternalism include safety regulations where goods have been banned or regulated because those who misuse them can cause a danger to themselves, including pharmaceuticals and recreational goods. If we consider the policy model $\max_{x \in C} U(x | \Theta)$ again, the whole notion behind these policies is that it is somehow possible to improve the outcome for the individual by reducing the choice set [C]. That is policies directed at making the decision-maker better off by being presented with fewer options. I will refer to policies that have as their primary goal to improve the welfare of an individual decision-maker by restricting the decision-maker's selected consumption bundle from other another that has been revealed to be inferior as being a paternalistic policy.

$$\theta \in \Theta$$

Accepting the premise of nudging has wide implications for measuring welfare. Fundamentally, behaviorally informed economic policy sets aside the long tradition of economics as a normative discipline. While a considerable amount of economics research is positive (being used to describe behavior), much of the work on policy analysis is normative, looking for policies that maximize social welfare. Indeed the foundation of behavioral economics is at complete and total odds with the foundations of welfare economics Koeszegi and Rabin (2007). It seems as if attempts to extend welfare theory to behavioral results must either set aside the empirical keystone of revealed preferences, or they must limit revealed preference within the confines of a single behavioral treatment or context. Either of these routes raises significant questions about the validity of application as a policy tool. For this reason, it is at once more important to make moral arguments for behavioral policy, and more important to recognize that policy is being made based solely on moral arguments. As such, accepting nudging requires a reinterpretation of welfare theory or at least to adjust our faith in these models. Along these lines, this thesis will argue for ways to justify nudging in a welfare context. Going forward this thesis will present one solution to this problem by re-evaluating nudging and its implications based on the insights presented in chapter 2 and 3.

PART 2. Redefinition and Debate of Chapter 1's Formulated Critique of Nudging

4 Chapter. Definition and Categorization of Nudging

This chapter will give a more precise definition of nudges, a complication to this is that the application and acceptability of nudges are usually discussed without determining or being sufficiently clear about whether certain requirements apply to the motives of the choice architects for an intervention to count as a nudge. Also, a series of conceptual details are underdetermined in the current literature. For instance, (Oliver, 2011) shows how it is unclear what the exact relation between nudges and incentives is. Furthermore, there seem to be difficulties in defining a strict line between nudges and other interventions? Several definitions and alterations of nudges have been suggested throughout the literature. In addition to the original definition presented in the beginning of chapter 1, (C. R. Sunstein, 2014) clarifies in a subsequent article that a nudge intervention must have the following characteristics:

- **Freedom of choice:** Nudges should always preserve the possibility for people to diverge from it deliberately.
- **Transparency:** It should never be used as a manipulative maneuver, and it should be possible for the public to review nudges in the same manner as other governmental actions.
- **Effectiveness:** If properly implemented, nudges should lead to protecting people against economic and physical harm.
- **Evidence:** Much like in any policy work, practitioners must rely on evidence to build a relevant Choice Architecture.
- **Testing:** When applied, nudges should be tested and prototyped, to anticipate and measure adverse consequences, allowing for iteration and avoiding extra expenses and efforts. It is only because of our frequent reliance on the automatic system that humans can be nudged. Without diving deeply into examples, here is a list of a few nudge techniques.
- **Default rules:** Setting people in a specific program, such as default double-sided printing in university printers or default choices of retirement plans in governmental policy.
- **Simplification:** To simplify information to avoid misunderstandings and facilitate navigation, and making complex navigation more intuitive.
- **Uses of social norms:** Emphasizing what most people are doing (e.g., "*most of your neighbors have insulated their attics*") is an effective nudge to engage people in a certain behavior.
- **Increases in ease and convenience:** The aim with this type of nudge is to make things easy: making healthy foods more visible is going to increase the possibility it gets picked.

- **Disclosure:** To make information accessible, for example, by demonstrating the environmental cost associated with energy on the energy bulb packs.
- **Warnings, graphic or otherwise:** Nudges can also be more explicit in describing the risk of some behavior, like the warnings that you can find on cigarette packaging.
- **Pre-commitment strategies:** To encourage people to engage in a specific course of action to reach their goals.
- **Reminders:** To avoid procrastination or forgetfulness, people can be nudged by being alerted of their upcoming obligations or commitments by email or text message.
- **Eliciting complementation intentions:** When asked about their intention to engage in a certain activity, people are more likely to do so, for example: “*Do you plan to vaccinate your child?*”.
- **Informing people of the nature and consequences of their own past choices:** Much like the disclosure nudge, however with emphasis on the individual’s data. E.g., “*your electricity consumption has been following a trend (...)*” (C. R. Sunstein & Reisch, 2014)

As stated in the literature review a nudge is an intervention aimed at redirecting an agent’s choices by slightly reorganizing the choice environment so that the interference in preferences is kept to a minimum and to make ‘suboptimal’ choices less convenient. In another sense, nudging is a policy intervention that reaches its objective by restricting the influence of rationality failures among its audience.

4.1 Significantly Changing Incentives

An important aspect of nudging is how to separate nudges from disclosure of information, and persuasion, is there anything novel about nudging? “*Provision of information is certainly a nudge, but it may or may not qualify as paternalistic*” (C. R. Sunstein, 2014, p.55) The inclusion in what counts as a nudge is confusing. It is clear that Thaler must have a particular kind of information in mind because they say: “*a nudge significantly alters the behavior of Humans, even though it would be ignored by Econs.*” (Sunstein & Thaler, 2008, p.8) There is nothing intrinsically irrational about selecting one action over another based on provided information. So what kind of information is available to the choice architect, if it is to be considered a nudge? One could assume to a more or lesser degree that the provided information (facts) is already known to the nudged, so the information, reserved for a choice architect, is to correct potential mistakes. A nudge can work as a reminder or make certain aspects of a choice more salient.

A paramount to nudging, is that they work without “*significantly changing incentives.*” (Sunstein & Thaler, 2008, p.5), this requirement is clarified by stating: “*In accordance with our*

definition, a nudge is any factor that significantly alters the behavior of Humans even though it would be ignored by Econs.” (Sunstein & Thaler, 2008, p.8).

The importance of this example becomes clear with an example from (Hansen & Jespersen, 2013). In lotteries, disproportionately large behavioral effects on 'Humans' may be obtained by making small, in principle insignificant changes in incentives that would not affect Econs. Say for instance you give an individual \$1 to complete survey. Another option is to each respondent that they are included in a lottery draw where the prize is an iPad. You are told that there are 500 respondents and the iPad is worth 500\$. Now, to a rational agent, this should make no difference since the expected utility in both scenarios would be \$1. However, for 'Humans', it makes a world of difference, since we tend to overestimate small probabilities (Burns, Chiu, & Wu, 2010). One may argue that Humans often perceive direct payment differently than they view lottery tickets. We usually regard the latter more similar to a gift, thereby activating norms of reciprocity, contrary to the former which we consider as impersonal payment.

The example shows that to be considered a nudge one cannot change the economic incentives, but rather restructure or rearrange the incentives. However, the 'left-digit-effect' of changing the price from \$1 to \$0.99 is a nudge and not a modification of the economic incentive Huber, Payne, and Puto (1982). This distinction is because the behavioral effect is not credibly derivable from an economic agent's utility function alone. The nudge is identified by the deviation from the predictions of the standard economic models. In this case, it is rather the 'salience effect' which causes the deviation. By understanding nudges as interventions which bring about a predictable irrational (when compared to standard economics) change. One can also add and remove choices, as long as they do not alter the behavior of an economic agent, this is because we suffer from "*dynamically inconsistent behavior*" (Sunstein & Thaler, 2008, p.41). One can present a decoy or more correctly make use of the 'asymmetric dominance' effect. In Huber et al. (1982) example people were faced with the choice between 2 different beers.

In (Huber et al., 1982)'s first example, people can choose between quality or quantity.

Type	Price six-pack	Brand rating 1-100
1	\$1.8	50
2	\$2.6	70

Now (Huber et al., 1982) adds a decoy. In this second example most people prefer option two because it seems more attractive, now that you get the same quality as the third type of beer, but at a lower cost. The third type of beer affects our choices because it adds possibilities for comparison, but its sole purpose is to increase sales of type 2

Type	Price six pack	Brand rating 1-100
1	\$1.8	50
2	\$2.6	70
3	\$3.00	70

The example portrays the 'asymmetric dominance effect.' This effect accounts for how adding an, in principle, the irrelevant option affects humans. The change should not influence an Econ because rational choices are independent of irrelevant alternatives (A. K. Sen, 2014). So adding or removing irrelevant options must count as nudges. The example shows that people, in general, do not have strictly defined pure preferences that tells us how much something is worth. We satisfice by focusing on the relative advantage of one thing over another and estimate value accordingly. People need a context to know whether they think the deal is good or not.

4.2 Provision of Information

Provision of information can be considered a nudge when it is directed against supposed mistakes. The choice architect can assume various biases like the nudged will forget a particular action, like forgetting the credit card in an ATM, but at the same time be perfectly aware of the task at hand. 'Post-completion error' (Byrne & Bovair, 1997). Provision of information can thus be considered nudging when it identifies an irrational reason for our behavior or lack thereof.

(Sunstein & Thaler, 2008) claims to help people to make better choices by influencing their decisions into choosing what they would regard as 'the all-things-considered best choice,' if given the time to reflect on their choices. It changes the Choice Architecture to assist decision-makers in making good choices (Kahneman, 2003; Starmer, 2000). Nudging is considered as part of soft paternalism because in general, paternalistic policies try to protect individuals from themselves. Libertarianism, on the other hand, focuses on autonomy, freedom of choice and the preservation of liberty. The combination libertarian paternalism can be understood as a "*relatively weak, soft and nonintrusive type of paternalism*" (Sunstein & Thaler, 2008, p.5). It works without explicitly preventing any options for the individual, thereby conducting influence with consent from the nudged. Following (Sunstein & Thaler, 2008), the main aim of modifications in the choice architecture is to create a choice context that allows humans to rely on their intuition instead of punishing decisions not based on reflective thinking. The ingenuity of Thaler and Sunstein is their presentation of libertarian paternalism to rationalize policy goals that appear to undermine the individual's preferences.

Libertarian paternalism is a philosophy that retains the essential element of welfare theory, which stipulates: benefits to others must compensate any restriction on choice (C. R. Sunstein & Thaler, 2003). Paternalistic policies are still permitted as long as they can be achieved with-

out changing the choice set. Behavioural interventions can be used to modify the context and framing of the decision rather than the choice set itself. The influence of nudging works in probabilistic terms. By creating a different frame or context, public officials hope to align the preferences of the individual with the policymaker. In relation to the model of individual welfare, the policymaker now selects $\theta \in \Theta$ (see last section in chapter 3), in hopes of improving the well-being of the individual.

This thesis will now return to the justifications for nudging. The first justification for nudging is related to their claim that humans are prone to “*dynamically inconsistent behavior*” (Sunstein & Thaler, 2008, p.41). This human characteristic shows itself in situations where context dependent preferences deviate from their deliberate preferences preceding the context. An example could be a self-imposed plan of cooking dinner and the later behavior of ordering pizza. Thaler and Sunstein do not claim that humans left defenseless against these flaws: they can apply internal self-control strategies, “*pizza once a month*” or mental restrictions on the use of money (C. R. Sunstein, 2000). Such a self-control strategy could be, for example, the use of different money jars for specific spending purposes (Sunstein & Thaler, 2008, p.50).

4.3 Health Policy Example

Health policy - like other policies – focuses on how to achieve the best health outcomes subject to budget constraints. They are often informed and directed by the discipline of economics (Backhouse & Backhouse, 2002). (Wansink & Hanks, 2013) have shown that individuals are significantly more likely to eat the first item in a buffet line. This behavior can be explained as ‘satisficing’ behavior rather than maximizing. Thus, if the first item in the line is good enough to meet the individual’s desire, they will select it, rather than exploring what they may be giving up later in the line. A nudge would place, the healthiest food item first, the location which requires the least effort to select. Similarly, the most conservative savings plan is the one that is adopted if the decision-maker does not actively request otherwise.

Significant changes can be obtained, simply by changing apparently trivial aspects of the choice environment. The way decisions are presented have significant effects and can, depending on circumstances, lead to important changes in behavior. Going back to the cafeteria example, a large body of evidence indicates that most of the time, we are paying very little attention to what we eat (Wansink & Van Ittersum, 2013). We are heavily influenced by cues in our ‘eating environment’, such as the size of the plate from which we eat food. In this regard, many simply consume everything on their plate, regardless of the amount. The action has also been labeled the ‘clean plate club’ (Wansink & Van Ittersum, 2013). As a result, larger plates usually increase our food consumption without us realizing it (Hollands et al., 2015). The findings suggest a straight forwards way of making it easy to eat less. Simply change packages, serving

items and tableware to a smaller equivalent. The key insight shows that vast opportunities for improving our health are possible if we recognize that our decisions rarely are deliberate and well-thought-out, but habitual, automatic and heavily influenced by the environment in which they are made.

4.4 Revisiting the Anti-Nudge Position

Thaler and Sunstein say that their recommendations are designed to “*make choosers better off, as judged by themselves*” (Sunstein & Thaler, 2008, pp. 5,10,80). The implication is that although the planner acts on his responsibility. Rules of thumb used in different contexts are unique and can vary widely. Thus, identifying these rules of thumb to alter decisions is a challenging and time-consuming endeavor, that requires intricate knowledge about decision processes. Thaler and Sunstein’s approach to nudging requires that choice architects can reconstruct the nudged’s latent preferences, even though observing actions cannot in a straight forward manner reveal these preferences, as they are “*dynamically inconsistent.*” Thaler and Sunstein claim that the findings in behavioral economics make paternalism unavoidable, the anti-nudge position is “*incoherent,*” a “*non-starter.*” In both (C. R. Sunstein & Thaler, 2003) and (Sunstein & Thaler, 2008) this claim is developed in relation to the following cafeteria example. The premise is that costumers’ choices between alternative food items are influenced by availability, the visibility of the elements displayed on the cafeteria counter. Knowing that some food is healthier than other, the cafeteria director has to choose the relative prominence of the different dishes (Sunstein & Thaler, 2008). Thaler and Sunstein considers two apparently reasonable strategies that a director might adopt: she could “*make the choices that she thinks would make the customers best off, all things considered,*” or “*she could give consumers what she thinks they would choose on their own*” (Sunstein & Thaler, 2008, p.2). Thaler and Sunstein tells us that the second option is “*what anti paternalists would favor,*” Thaler and Sunstein want to show why option 1 is the sensible choice and freedom preserving. The first argument begins with the observation that people exhibit “*dynamically inconsistent behavior.*” Thaler and Sunstein believe that individuals:

“lack well-formed preferences, in the sense of preferences that are firmly held and preexist the director’s own choices about how to order the relevant items. If the arrangement of the alternatives has a significant effect on the selections of the customers make, then their true ‘preferences’ do not formally exist” (C. R. Sunstein & Thaler, 2003, p.6).

The viability of the anti-nudge position is reduced, by indicating that choice architecture exists without an architect and by implication that nudges may exist without intentionality. It appears that individuals are nudged by the choice architecture, albeit not necessarily planned with any deliberate intent.

After this conclusion, Thaler and Sunstein can argue, that there is no such thing as a 'neutral design' even when the nudge and its effects are accidental. They suggest that though no one may have intended to nudge us toward particular ends. Individuals are nevertheless always nudged towards one option, rather than another in a predictable and consequential way. Thaler and Sunstein conclude, despite being paternalistic, that a choice architect should design the nudge toward what the architect thinks would make people better off, this is the only reasonable option for a well-intentioned choice architect. However, this is acceptable because the context of choice always influences our choices, whether individuals like it or not. "*The anti-nudge position is unhelpful - a literal non-starter*" (Sunstein & Thaler, 2008, p.11).

4.5 As Judged by Themselves

There may be severe obstacles to determining what people judge to be in their interests, especially in the cases where they have not expressed or thought much about their preferences. Even if one asks for their opinion, there may be problems caused by the plasticity of preferences which is affected by how questions are presented and the ordering of possible answers, to the extent that makes concrete conclusions difficult (C. Sunstein R, 2017). For instance, when asking citizens about when and moreover, for what purpose they deem state intervention acceptable. Answers may depend on; the level of abstractness of questions and what is considered socially desirable. All these factors influence the answer to such an extent that even reflected answers might become inconsistent (C. Sunstein R, 2017; C. R. Sunstein, 2016a, 2017). If these answers are inconclusive choice architects are left to fall back on the "*as judged by themselves*" rule. (C. R. Sunstein, 2016b) admits that it can be hard for a choice architect to determine when the nudged individual is better off. They state that "*as judged by themselves*" should be understood as: "*The judgement of the person who is no longer in the grip of addiction or some other problem that is undermining the capacity for self-control*" (C. R. Sunstein, 2016b, p.47). Furthermore, they admit that there is a fine line between determining whether something is a self-control problem or a legitimate short pleasure.

Thaler and Sunstein does however not add any concrete suggestions and with their suggestive comment "*When there is a divergence, choice architects should follow people's reflective judgments*" (C. R. Sunstein, 2016b, p.48), their recommendations to respect 'legitimate' pleasures seem to be neglected. Thaler and Sunstein are aware of the difficulty in determining appropriate defaults One answer from Thaler and Sunstein seems to come when they state that

nudges should guide individuals: “*decisions that would change if they had complete information, unlimited cognitive abilities, and no lack of willpower.*” (Sunstein & Thaler, 2008, p.4). Moreover, Thaler and Sunstein refers to decisions deviating from this ideal as “*inferior decisions in terms of their own welfare*” (Sunstein & Thaler, 2008, p.4).

When it comes to an understanding of human actions, rational choice theory and nudging seems remarkably similar. For Thaler and Sunstein, the criterion of individual well-being is given by the preferences an individual would have revealed unaffected by bounded rationality. The choice architect main task is to somehow reconstruct individuals’ underlying latent preferences by testing their preferences when uninterrupted by reasoning imperfections. Implicit in this argument, one must assume that latent preferences are context-independent. Thaler and Sunstein seems to claim that, because revealed preferences are context dependent choice architects are justified in applying nudging, to guide people towards their latent preferences. Furthermore, they play on a softer requirement: if in any case people are affected by context, then choice architects must make sure that the situation does not guide people towards unsustainable behavior. Thaler and Sunstein believes that, in any event, individuals are affected by choice architecture, so choice architects might as well direct people towards what is best according to what is rational. As a pragmatic middle road Thaler and Sunstein ends up suggesting that choice architects should analyze the number of opt-out under different rules and ask people about their reasons for opting out, to improve the default.

4.6 Preferences

Going back to the quote “...*true ‘preferences’ do not formally exist*” (C. R. Sunstein & Thaler, 2003, p.6), it seems sensible to expect that Thaler and Sunstein are using ‘preference’ in the sense that it is used in conventional economic theory, as presented in chapter 3. In conventional economic theory preferences are consistently revealed in an individual’s actions, such as buying particular objects. Thaler and Sunstein assert that the people do not exercise these well-defined ‘true’ preferences among items. “*In many cases, individuals make pretty bad decisions - decisions that they would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control*” (C. R. Sunstein & Thaler, 2003, p.5). This view of preferences will be labeled latent preferences and refer to the preferences that a person would have acted on if unaffected by a context. Based on Thaler and Sunstein’s evaluation of the “*as judged by themselves*” rule, it seems as if choice architects are obliged to guide the individual towards the ideals inherent in rational choice theory. The rationale for latent preferences seems to hinge on the approach of behavioral economics in general. (Sunstein & Thaler, 2008, p.7) assert, that using rational choice theory as a basis for public policy-making, fails to take into account how context heavily influences individual decision-making and it could, therefore, be optimized. The disciplines of behavioral economics, cognitive, and social psychology have consistently conducted experiments disqualifying these

idealized axioms as the correct guideline for governing behavior (Kahneman, 2003). People are thus incapable of acting according to rational choice theory because they are led astray by their context.

The choice architect should still guide individual behavior towards the ideals inherent in rational choice theory, even though these true preferences cannot be observed.

Nudging treats cases in which an individual's choices depend on 'irrelevant' properties like framing as errors. Mistakes are defined relative to the latent preferences that the individual would have revealed if not subject to bounded rationality. Latent preferences are assumed to satisfy traditional principles of rational consistency, in particular, context independence. The satisfaction of latent preferences is taken as the normative criterion in nudge interventions. By latent preferences, one can understand tastes and knowledge formed within the minds of individuals. These preferences do not correspond directly with properties of the external world. Thaler and Sunstein appeal to this understanding of preferences when they repeatedly mention their goal is to make people better off "*as judged by themselves.*" If public planners treat welfare as a subjective concept, then nudging can be defended as a way of correcting misconceptions in an individual's reasoning while respecting his understanding of welfare. The tricky part is, if latent preferences is a subjective concept, one must assume that individuals can somehow access an unbiased reasoning that generates these latent preferences. In a strict sense, the empirical foundation for welfare analysis is undermined because of revealed preferences is an unsatisfactory method to measure welfare change according to behavioral economics.

This limitation in nudging is perhaps a byproduct of the structure in the standard theory of rational choice. Rational choice theory is a purely descriptive theory and hence do not try to explain the reasoning by which individuals construct their preferences (A. Sen, 1973). The issue with nudging is its attempt to guide the welfare of individuals "*as judged by themselves.*" This criterion is to some extent a black box choice architects cannot access. The best way, going forward is to minimize the discrepancy between individuals stated and revealed preferences.

Several studies have tried to reveal biases by testing choices "*no one could possibly want*" (Akerlof & Shiller, 2015, p.xii). Through experiments tests and questionnaires, one can approximate individuals latent preferences to a greater precision with nudges, than with rational choice models. At least this is the assumption in behavioral economics. If biases have been identified satisfactorily then, presumably, each nudge is somewhat controllable, in the sense that a policymaker can use it to guide individuals to choices that are preferred from the policymaker's perspective and likely to be preferred by the individual, while leaving room for other options. Moreover, the nudge itself must be relatively inexpensive to implement so that there is little welfare loss. Thaler and Sunstein ask:

"if subtle changes in the behavioral and decision-making context lead us astray from our own best interest. Why not use these insights into how and why this

happens to gently ‘nudge’ us into improving our health, wealth, and happiness”
(Sunstein & Thaler, 2008, p.7).

Accepting this premise the issues with nudging becomes tied to the choice architect. How can a choice architect legitimately influence the choices and behavior of others? It seemingly follows that individuals are nudged in perpetuity, regardless of anyone intended it so. Choice architecture is an inevitable feature of any choice. Accepting the previous conclusions, it seems reasonable to prefer nudging according to the principles laid out by Thaler and Sunstein.

Thaler and Sunstein state, a choice architect, whether she likes it or not, cannot avoid influencing the decisions and behaviors in the context she is responsible for organizing (Sunstein & Thaler, 2008, p.10). They even go further and claims that a choice architect’s capacity to nudge people in a healthier direction imposes a moral obligation on the choice architect to do so. Namely because, one can always reject the behavior that a given nudge is devised to promote (Sunstein & Thaler, 2008).

Nudging is presented as a ‘non-starter’ because it avoids limiting the original choices set, and changing existing economic incentives. The bottom line is, independent of the choice architects actions; individuals remain free to choose otherwise. The final argument asserts, that if guided by libertarian paternalism and the ‘Rawlsian publicity principle. If following these principles then the choice architect has satisfactorily tried to meet relevant political and normative concerns. It is important that a nudge can be defended in public. ¹

The addition of nudging to the public policy toolbox provides social planners with the seemingly most nonintrusive means to behavioral change that simultaneously takes seriously, that provision of information is not always as effective at changing behavior as believed (C. R. Sunstein, 2016b). The individual’s realized welfare for any outcome, should not be altered by the nudge, though the perception of utility in relation to other choices might be. According to Thaler and Sunstein, it seems that if one understands the extent to which biases influence us, one would be pro-nudging. Taking another view it may also seem as if people are unable to have no serious complaints about being nudged for one’s benefit, thus nudging is nonintrusive in all aspects. It is not forcing citizens nor limiting their freedom to choose otherwise (Dolan et al., 2012).

In the neoclassical account of decision-making and behavior, there is no room for nudging. The neoclassical ideas bear little resemblance to what behavioral economics has revealed about our actual everyday decision-making and behavior. This behavior is often influenced in

¹ The ‘Rawlsian publicity principle’ holds that, when justifying the exercise of political power, lawmakers ought to appeal exclusively to reasons that all “*citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason*” (Gosseries & Parr, 2017).

systematic ways by subtle, seemingly insignificant changes in the decision context. Nudging is claimed to have a sizable impact on behavior that is unintuitive to our self and especially to those more embedded in a rational choice doctrine. It is precisely this intention-action gap that motivates and justifies nudging. It is often claimed that nudges are so innocuous that the individual would not mind the nudge even if it were revealed (Jung & Data, 2016).

Inspired by (Hansen & Jespersen, 2013), Nudging is a broad term to describe deliberate and predictable methods of changing people's behavior by modifying the factors in the physical and social context in which individuals act. Nudges are justified because they target cognitive boundaries, biases, routines, and habits in individual and social decision-making contexts, posing barriers for people to act accordingly to their self-declared interests. Nudging influence people by incorporating knowledge of the boundaries, biases, routines, and habits to inform the design of the intervention.

The use of nudging implies that one particular choice is selected unless the user prefers otherwise. The perfect nudge is suggestive. In cases where nudging does not engage with individuals reflective capacities, the burden of proof is significant. The nudge can still be considered suggestive, but since the implicated choice is likely to be the de facto option chosen.

With a great deal of insight Thaler and Sunstein state that: "*choice architecture can claim to be rooted in an incompletely theorized agreement - an agreement from people who do not agree on the foundational questions in moral and political philosophy, or who do not know what they think about those questions*" (C. R. Sunstein, 2016b, p.52).

5 Chapter. Intentions and Ethics

5.1 Manipulation

Thaler and Sunstein are certainly right in saying that the context for our decision influences our choices. However, they do seem to convey vague and insufficient information about the issue of intentionality. The relationship between autonomy and influence is the first point in this chapter. The challenge lies in identifying to what degree people are free to choose.

Much of the discussion about nudge in philosophy has focused on a normative assessment of Libertarian Paternalism, such as asking whether Nudges are Paternalistic, or manipulative, or whether they erode or improve a decision-maker's capacity to make informed decisions (Grüne-Yanoff, 2012; Hausman & Welch, 2010; Wilkinson, 2013). The normative aspects of this discussion have received considerable attention.

A persistent and influential debate has been the claim that nudging works by 'manipulating people's choices' (Bovens, 2009). Under certain circumstances, Thaler and Sunstein themselves subscribe to this view (Sunstein & Thaler, 2008). Another criticism often derives that nudging is just paternalism in disguise and return to present day behaviorism (Burgess, 2012). It seems as if the anti-nudge position, might not be a 'non-starter.' The characterization of nudging as the manipulation of choice, and the policy recommendations that results from this characterization, depends on the theory of agency that one subscribes to, and how this attaches to ethical considerations and normative responsibilities. In particular, it is a problem that much of the criticism are rooted in theories of agency and ethical considerations quite different from the theory and concepts underpinning the nudge approach to behavioral change.

5.2 Evaluating Nudging

In Chicago, a nudge intervention focused on stimulating drivers to slow down by painting white stripes on the roads. These stripes were initially evenly spaced but get closer together as drivers reach a dangerous curve. This design gives the sensation that driving speed is increasing and intended to triggers the driver's natural instinct to slow down. (Sunstein & Thaler, 2008, p.39). This example shows a positive relationship between the involvement of cognitive processes in the decision-making process and the effectiveness of nudging.

Nudging interventions can undoubtedly improve and simplify the human decision-making process by creating or reshaping the cognitive biases generated by 'System 1' biases, which determines automatic responses or generates the cues that underlie reflective decisions (Amir & Lobel, 2008). In particular, this example shows how a nudge can be recommended, but one should also note, how this type of visual illusions is almost irresistible. A far stretched implica-

tion could be that people, in the long run, feel encouraged to disregard their sensation of speed, after becoming aware of the visual illusion caused by the stripes, i.e., leading to an increase in reckless driving.

Interventions which effectively changes our behavior through non-deliberative processes, also involve ethical issues, as noted earlier, because they involve altering behavior through mechanisms of which people are not obviously aware. Nudging like the Chicago example are thus more acceptable if there is good evidence that they will be effective in tackling harmful effects on the population.

5.3 Transparency

Professor Luc Bovens, London School of Economics and Political Science, suggested two distinctions of transparency, with specific relevance to nudging. Transparency can either require the choice architect to tell people that they will be nudged, or ensure in the design of the nudge that an insightful person can identify for themselves that they are nudged in the given context (Bovens, 2009, p.12). Bovens emphasizes that the latter (weaker) form distinguish nudges from subliminal messaging, it is thus not manipulative in the sense of an influence people are incapable of recognizing. Another reason why Bovens suggests this weaker form is that he believes “*nudges work better in the dark.*” He asserts that governments are not required to explain that an intervention has been implemented because it might limit the effectiveness of the intervention. Nudges are deemed acceptable provided that it is reasonably clear, that a nudge has been devised, regardless whether individuals notice it in practice. It seems as if the line dividing an intervention between possible to discern but seldom will and impossible to discern, is, imprecise at best. At the same time, one should note, that this version of transparency is all that is required by businesses (Science et al., 2011).

It seems dishonest to appeal to this freedom in a practical context when nudging is applied exactly in contexts where individuals fail to make use of reflective thinking (Hansen & Jespersen, 2013). The claim immediately appears at odds with intuition. The Danish behavioral scientist Pelle Guldberg Hansen and Ph.D. student Andreas Maaløe Jespersen argues that in the normative justification for nudges, one has to consider the issue of intentionality and, by extension, agency. Hence they say

“There seems to be a clear and important distinction to be made between a given context that accidentally influences behavior in a predictable way, and someone – a choice architect – intentionally trying to alter behavior by fiddling with such contexts” (Hansen & Jespersen, 2013, p.10).

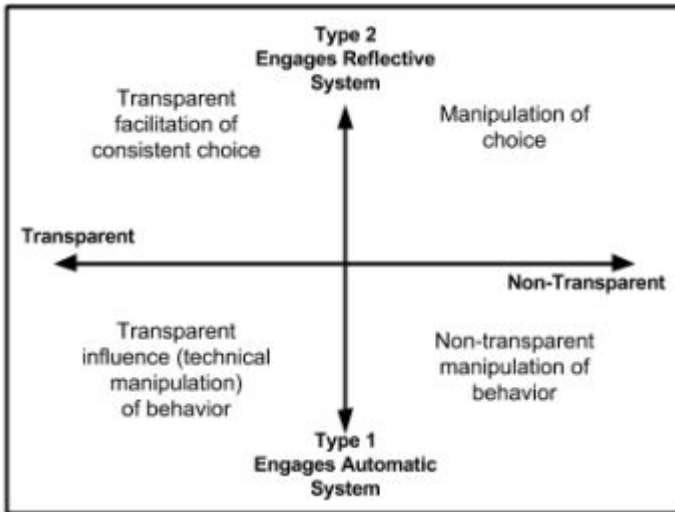


Figure 2: Matrix of Transparency

Intentionality is a precondition of normative evaluation. Unfortunately, as Hansen and Jespersen also point out, this absence of responsibility is often indicated when stating that nudges are everywhere. Such a move blurs the identification of responsibility.

Because nudges consist in a vast array of different interventions, (Hansen & Jespersen, 2013) have developed a matrix to determine a nudge's level of transparency and to what extent it engages with individuals' automatic and reflective decision-making systems. According to Hansen's framework, nudges range from most manipulative (Non-transparent Type 1 nudges) to least manipulative (Transparent Type 2 nudges). Non-transparent Type 1 nudges are those in which individuals are unaware they are being nudged and act automatically without thinking. This prevents one's ability to 'opt out' from the target behavior and thus gives nudge practitioners control over individuals' decision-making (Hansen & Jespersen, 2013). This nudge should be avoided unless there are exceptional reasons for its implementation. Conversely, Transparent Type 2 nudges alert individuals of the nudge and engage their reflective systems, allowing individuals to easily 'opt out.' This reassures that the target behavior is one that individuals intend to do, and because the reflective system is activated, the target behavior has a higher chance of being repeated in other contexts. The researchers recommend nudge practitioners strive for this type of nudge above the others both for ethical reasons and for its potential for teaching healthy and sustainable behavior that will translate into other situations. This however greatly reduces the applicability of nudging. Going along with (Hansen & Jespersen, 2013) a nudge might best be understood as the deliberate attempt from a choice architect to influence choice. It is still acknowledged that the settings of any given decision context may accidentally change choices and behavior in predictable ways as well. However, nudging should only be applied when the context is intentionally altered towards a specific behavior change.

5.4 Nudging an Intentional Act

Nudging then seems to make the most sense as an intentional act, especially when discussing manipulation. However characterizing nudges simply as manipulation, is too simplistic. The psychological dual process theory that underpins behavioral economics, used by Thaler and Sunstein distinguishes between automatic processes/behaviors on the one hand and deliberate choices on the other. Nudging always influences 'System 1,' since all sensory input is dealt with by this system and only sometimes does nudging effect our reflective system (Hansen & Jespersen, 2013). The theoretical implication of this is that nudging only sometimes the actions individuals perceive as choices, and we are in this sense not engaging with the nudge, but merely following its suggestions.

5.5 Manipulation and its Context

Nudge has been referred to by (Sunstein & Thaler, 2008) as libertarian paternalism as this method organizes choices without force. In principle, the recipient remains free not to perform the action to which they are being compelled. While in medicine, paternalism has a disapproving connotation, this may not be the case in the realm of public policy. As pointed out by Israeli philosopher Cohen “ *the meaning of paternalism is quite dissimilar in public policy or legislation and interpersonal relations*” (Cohen, 2013, p.13). With this distinction, the moral evaluation of nudges becomes necessary. If it is designed to pick off efficacious influences that preserve freedom of choice, yet bypass the deliberate capacities of those influenced. Although some nudges preserve freedom of choice, it is necessary to elaborate what this preservation entails. Freedom of choice requires caring for the control an individual has with their assessments and choices (Saghai, 2013). For (Saghai, 2013), a crucial condition for the ethical acceptability of nudges is to maintain the capacity for deliberation in individuals. This ideal for autonomous decision-making as expressed by Saghai is one of the main venues for criticizing nudges. According to (Hansen & Jespersen, 2013), manipulation occurs when there is a willful deception that remains concealed. Certain nudges do bypass the deliberative capacities of those influenced. With the above understanding of manipulation, these authors believe that nudging affects individual perceptions, choices, and behaviors through devious or abusive tactics.

According to (Dubov, 2015) persuasion is the process of using arguments to convey information potentially leading to change in the behavior or opinions of another person. If a person is persuaded through means of manipulations, then the individual may falsely believe that they were persuaded by the information relevant to the topic. When in fact, their deliberative capacities are circumvented by withholding critical information or similar methods aimed at distorting the reality of the situation. “*Manipulative strategies are designed to create an illusion of free*

choice to lure the victim into a decision that would probably be resisted under normal circumstances” (Dubov, 2015, p.21). The difference between manipulation and persuasion is necessary, and it provides an ethical justification for the choice of particular techniques as opposed to others. Manipulative techniques can be used such as the ‘foot-in-the-door²’ technique, and the ‘commitment effect³’ describes how people, when committed to a decision, become resistant to changing their minds. These authors ignore that removing the biases that give individuals an impulsion towards ends that are in contradiction with what they think to be in their best interests also counts as persuasion because the choice architect is attempting to improve the recipient’s opportunity to make a rational choice. More positively, nudges with a sound evidence base designed through honest communication, with care not to mislead through exaggeration, can hardly be counted as manipulation.

5.6 Blood Donation

A good example of an acceptable nudge is the Swedish Blood donation case. (Buyx, 2009)states that “*Richard Titmuss’ influential work «The Gift Relationship» (1971) resulted in most countries in the developed world abandoned paid blood donations and have either fully altruistic donation systems or mixed systems with altruistic donation*” (Buyx, 2009, p.1). Blood donation programs increased their efforts to pay more attention to the motives, the choice architecture, and the settings for donations. An experiment in Stockholm, Sweden recently began to send a text message when a donors blood has been accepted, and a second message when received by an individual. This nudge cannot be characterized as manipulative. The problem is rather that we cannot know for sure whether this initiative will have a positive effect on donation rates. Several conclusion is possible based on the delay between donation and use, which could lead donors to conclude that the supply of blood is much higher/lower than they previously believed, thus greatly influencing individual motivation.

5.7 Self-Interested Nudges

To alter the behavior of individuals, one must occupy them with new norms, to install a reliable behavior change. The goal is to reduce the potential for conflict and rejection between the population, and directing the population in ways that: “*do not attempt, at least not primarily or in a fundamental way, to make use of a relationship of obedience (. . .)*” (Foucault, 2007, p.65).

²is a compliance tactic that involves getting a person to agree to a large request by first setting them up by having that person agree to a modest request

³Refers to a human behavior pattern in which an individual or group—when faced with increasingly negative outcomes from some decision, action, or investment—continues the same behavior rather than alter course.

To the same extent Thaler and Sunstein state: “*We elaborate a form of paternalism, libertarian in spirit, that should be acceptable to those who are firmly committed to freedom of choice on grounds of either autonomy or welfare*” (Sunstein & Thaler, 2008, p.3).

Nudges operate by providing information overtly and covertly at the same time to amplify the possibility of informed decisions. “*In effect, what defines a relationship of power is that it is a mode of action which does not act directly and immediately on others. Instead, it acts upon their actions: an action upon an action, on existing actions or on those which may arise in the present or the future*” (Foucault, 1982, p.789). At the same time, this micromanagement and behavioral feedback make monitoring of behavior increasingly insightful. Thus, by applying nudging, marketers can alter the very conduct of consumers, by promoting the consumption norms which they benefit from. The very exercise of setting up nudges provides marketers with an increasingly effective tool to influence consumers. This reveals the intertwined relationship between knowledge and power. An individual does not hold power. Instead, power is a contemporary agreement; it is a ‘regime of truth’ that pervades society, and which is in constant flux and negotiation.

Absolute truths that can be discovered and accepted does not exist. Truth’s existence is embedded in the methods and rules according to which science separate the true and false knowledge:

“Conduct is one of the best aids for coming to terms with the specificity of power relations. For to “conduct” is at the same time to lead others and a way of behaving within a more or less open field of possibilities. The exercise of power consists in guiding the possibility of conduct and putting in order the possible outcome” (Foucault, 1982, p.789).

The exposition of nudging aims to move individuals in a certain direction without restraining their freedom of choice. As a result, ‘nudging for good’ seems to promise a management of people targeting psychological triggers underlying human behavior and choice to produce healthier, safer lives, and increased welfare. The clear implication is that extreme or highly unwelcome defaults are less likely to stick. It follows that for example green defaults that are perceived as foolish, wrong, harmful, expensive, or as the imposition of some high-minded environmentalist elite, may well be rejected by many consumers. If people are highly skeptical about public officials, and if such officials set the default, they may reject it for that reason.

5.8 Responsiveness to Nudging

A more puzzling and somewhat troubling implication, based on the population with lower incomes, is that default rules may be more sticky for low-income workers than for their higher-

earning counterparts (C. R. Sunstein, 2016b). One reason may be that low-income workers have a great deal to worry about, and so are less likely to take the trouble to think through and to alter the default rule. An 'effort tax' may seem especially high for, and have an especially large adverse effect on, people who are already facing a large number of decisions and costs (C. R. Sunstein & Reisch, 2014, p.145).

In the literature authors judgments about nudges makes sense and line up fairly well with the conclusions that follow from the extended analysis. People do not oppose nudges as such. So long as they believe that the end is legitimate and important, and likely to be favored by choosers themselves, most people will support nudges in its direction. By contrast, people dislike those nudges that (a) promote what they see as illicit goals or (b) are perceived as inconsistent with either the interests or values of most choosers (C. R. Sunstein, 2016a). Avoiding the trap of abstractions, people agree that the ethics of nudging depends on the concrete nudge in question. In addition, people tend to prefer nudges that target and improve deliberative processes to those that target unconscious or subconscious processes, and they may react against the latter, but they people do not reject the latter and will often approve of them as well (C. R. Sunstein, 2016a). When the political valence of nudging is clear, their evaluation of nudges very much turns on that valence: People on the political right tend to approve of nudges that move people in the direction that they favor, and people on the political left show the same tendency (C. R. Sunstein, 2016a, 2017). This finding reinforces the view that in many cases, people are more concerned about the ends nudged towards, rather than of nudging as such (C. R. Sunstein, 2016b, p.200) and most people would endorse nudging if asked (C. R. Sunstein, 2016a).

Nudges that focuses citizens towards responsible behavior seems both legitimate and welfare maximizing for society because it is exercised to support rather than to abuse our cognitive deficiencies: *"Transparency about nudging will not, in general, reduce the effectiveness of nudges, because most nudges are already transparent"* (C. R. Sunstein, 2016b, p.1). The crucial question to answer here is what extent are people are responsive to nudges. Nudging is said to allow individuals to choose freely, but also always implies that a certain behavioral option is made more attractive and intuitive than others. Given that nudges are easily resistible, it distinguishes itself from more coercive implementations. Both approaches influence the individual but with a widely different intensity. Nudging enables people to use their freedom and is designed to promote positive effects for health, self-efficacy, and safety. As a consequence, it is the mind in specific contexts that need to be bypassed or 'boosted' to make responsible choices.

5.9 Can Nudging be Resisted?

Foucault emphasizes that power be often presented as an expression of goodwill, as mentioned in the methodology. Foucault argues that this form of control is often more efficient because it allows for a more economical and predictable form of power. Thaler and Sunstein argue along

the same lines, by suggesting that the bottom-up power of nudging is; *“far less expensive, and less intrusive than the command-and-control approaches”* (Sunstein & Thaler, 2008, p.200). Nudges have different effects depending on the personality traits and interests of the specific individual in question. Some will find pleasure in figuring out choices by themselves, other will find great relief in defaults, and these preferences can change depending on the topic, health care, tax returns, etc (C. Sunstein R, 2017). In conclusion, nudges work as a form of standardizing power, rather than a top-down form of persuasion. People tend to accept nudges that they think are consistent with *“the interest or values of most choosers”* (C. R. Sunstein, 2016b, p.7), finding the interest of ‘most choosers’ is, however, a practical and changing issue, which is paramount to the success of nudges. The acceptance of nudging is necessary before policymakers can use apply it to alter peoples choices. (Hansen & Jespersen, 2013) argue that the problem of nudging is that it is extremely challenging for citizens to reconstruct either the intention or the means for why they changed their behavior. People fail to choose optimally, either because of internal biases or because of social norms dictate it. People develop all kinds of harmful habits by copying the behavior of others, despite the behavior being damaging to their health or safety. As discussed in this thesis, in order for a nudge to be manipulative, it must be undetectable by its recipient, and direct the recipient towards questionable ends. If the recipient is unaware of the nudge, her right to choose and learn from that choice has been removed by those in a powerful position, and the value of human autonomy has been interrupted.

6 Chapter. Challenging implications as to scientific findings and implementation in public policy

6.1 Welfare

Behavioral economics suggests that choice architecture is a means through which to achieve previously defined ends best, while remaining neutral between such ends. If the ends themselves cannot claim to be value neutral, then nudging towards a specific value must surely push the recipient toward that same value. It must be recognized that programs such as 'Save More Tomorrow' has a clear normative agenda with underlying assumptions that might run contrary to the underlying principle of value neutralism. In a meta sense, the idea that the goal of policymaking is to promote rational decisions simultaneously assumes that rationality, including preference consistency, is in itself a worthy cause. Whether this is right or wrong is not within the scope of this thesis. Undoubtedly some people would find the immediate satisfaction and gratification of greater value, than future savings and long-term health, while taking into account the risks these choices would bring about. Thaler and Sunstein do not deny the possibility. They do, however, find their tool nudging liberty preserving, if it helps individuals, who, because of their biases find themselves unable to commit to their desired level of retirement savings. On the other hand, they do not deny people who have thought their choices through to do as they see fit. To make a clearer example, individuals who participate in extreme sports are well aware of its accompanying risks, i.e., they do not mentally discount the risks, but find intrinsic value in this particular lifestyle (Loewenstein, 1999). If we look beyond the individual and focus on the economy, in general, we can see that government intervention aimed at helping consumers with their lending, saving, and spending, then some corporations will lose money, as these biases are a great source of revenue (Egan, 2005; Harris & Albin, 2006)

6.1.1 Lasting?

The difficulty facing proponents of nudging is that this approach works more efficiently in situations where the individual struggles to understand the situation (Sunstein & Thaler, 2008, p.247). While at the same time people tend to express greater support for nudges that appeal to our rational faculties, than nudges targeting routines and automatic processing (C. Sunstein R, 2017). In general, such decisions might require vast technical knowledge to remove negatively biasing components. In these situations, it might also be unclear what the individual's preferred judgment is, given that the subject of the nudge may struggle to consider the potential impact of their choices (Mills, 2013). As (Sunstein & Thaler, 2008) points out, people are most likely to need nudges for decisions that are difficult, complex, and infrequent combined with poor

feedback and few opportunities for learning.

There are of course many unresolved issues in applying nudge-like interventions. What is the duration of the various effects? On the face of it, the effects of, say, priming appear fleeting, they "*the effects of priming may only last for a short while after exposure to the prime*" (Dolan et al., 2012, p.75). This uncertainty does not mean, however, that their impact is insignificant, since the behavior and decision may have been changed in that interval: the priming effect may have led to someone making a commitment that translates into longer-lasting change. Thus, effects such as priming may be thought of as ‘triggers.’ Others may be ‘self-sustaining’ effects: once enacted, their mode of operation supports continuity. For example, the use of defaults is based on the status quo bias, which encourages stability and minimum effort over time.

6.1.2 Practical Evidence

Another problem is the relatively limited practical evidence about how the effects might habituate within individuals over time (Dolan et al., 2012). Success will probably depend on whether the individual is happy with the result – in other words, whether they also find the change desirable, the reinforcement that follows it. The most acceptable interventions will certainly be those that seek to change minds alongside changing contexts. Smokers who are trying to quit deliberately avoid some of the primes that encourage their smoking, such as the habit of having a cigarette with a drink. Effects that direct people away from smoking are assumed likely to be welcomed rather than consciously resisted. The effect might even be reinforced by the sense of feeling good after resisting the temptation or seeing that their efforts have worked.

The *UK House of Lords Science and Technology Committee* noted that behavioral interventions are difficult to assess.

“Our impression that there is relatively little evidence of the effectiveness of particular behaviour change interventions at a population level has been reinforced by how few substantial responses we received following our request for examples of successful interventions” (Science et al., 2011, p.18).

For once the evidence and methods used in behavioral economics are highly debated. There is a lack in how a certain intervention (choice architecture) works and to which extent it will be effective in a specific environment. The experiments are often described as evidence based. However, it seems unclear whether the described change in behavior can be assigned to one attribute or whether it was caused by another effect, i.e., loss aversion or the recommendation effect. Nudges do not seek to reduce rationality per se; it aims to reduce irrationality by intercepting sub-optimal decisions. The extent to which laboratory settings can capture the speci-

ficiencies of social situations remains questionable (Rostain, 2000, pp.984-8). Even if several tests now are conducted as ‘field experiments’ (James & Moseley, 2014). It is still questionable whether these ‘real life’ field experiments can overcome challenges regarding, scale, time and contamination effects. These issues of predicting the outcomes from nudging are only escalating when combining several intervention methods compared to singular nudges tested in the isolation of a lab. While the political limits to nudging resemble those of other policy areas, it is particularly noteworthy and ironic that an approach is claiming to be dealing with reasoning failure and bounded rationality show so little self-awareness of its limitations.

“Though nudge-economics remains seductive, what once seemed like a panacea has come to look a bit more like a series of sticking plasters” (Adams, 2014, p.2).

Despite the current enthusiasm amongst governments and policy-makers for behavioral approaches, one must be aware of the potential problems with behavioral sciences capacity to formulate public policy, many of these remain underexplored or are unsatisfactorily answered. The existence of empirical evidence for people’s deviations in rational choice context, including evidence for biases and an unwarranted reliance on heuristics, is often used as the justification for ‘nudge’ techniques.

6.2 Implementation Problems

6.2.1 A Taxonomy of Nudges

To accommodate the use of nudges, several authors have proposed various taxonomies to categorize the different kinds of nudging interventions coherently. (Hansen & Jespersen, 2013) proposed that nudges can be separated into an unconscious and conscious category, and transparent or nontransparent (Hansen & Jespersen, 2013). Their model is useful as a quick guideline to determine whether a nudge can be deemed manipulative. Several other taxonomies are available (Hollands et al., 2015; Ly & Soman, 2013; Science et al., 2011). A wide variety of policies affects the way people behave. (Science et al., 2011) Suggests a possible taxonomy with including examples of different types of intervention. Their model takes into account that most public policies intended to create behavioral change. Their taxonomy takes this into account and specifies how nudges can be considered as a particular type of non-regulatory interventions.

In line with the above table, nudges are to be understood as a subset of, non-regulatory interventions. The strength of (Science et al., 2011)’s model, is that it makes clear, that nudges are to be understood as a compliment to a specific intervention, they cannot stand alone. Nudges are designed to save people from the cognitive effort of making a choice, and so do not persuade through argument, unlike media campaigns and provision of information.

TABLE 1

Table of interventions

	Regulation of the individual		Fiscal measures directed at the individual		Non-regulatory and non-fiscal measures with relation to the individual					
	Eliminate choice	Restrict choice	Fiscal disincentives	Fiscal incentives	Non-fiscal incentives and disincentives	Persuasion	Provision of information	Changes to physical environment	Changes to the default policy	Use of social norms and salience
Choice Architecture ("Nudges")										
Interventions category	Guide and enable choice									
	Examples of policy interventions	Prohibiting goods or services e.g. banning certain drugs	Restricting the options available to individuals e.g. outlawing smoking in public places	Fiscal policies to make behaviours more costly e.g. taxation on cigarettes or congestion charging in towns and cities	Fiscal policies to make behaviours financially beneficial e.g. tax breaks on the purchase of bicycles or paying individuals to recycle	Policies which reward or penalise certain behaviours e.g. time off work to volunteer	Persuading individuals using argument e.g. GPs persuading people to drink less, counselling services or marketing campaigns	Providing information in e.g. leaflets showing the carbon usage of household appliances <i>*Regulation to require businesses to use front of pack nutritional labelling, or restaurants to provide calorific information on menus</i>	Altering the environment e.g. traffic calming measures or designing buildings with fewer lifts <i>*Regulation to require businesses to remove confectionery from checkouts, or the restriction of advertising of unhealthy products</i>	Changing the default option e.g. requiring people to opt out of rather than opt in to organ donation or providing salad as the default side dish

Note: * Demonstrates how regulation of businesses might be used to guide the choice of individuals, thus distinguishing it from regulation which restricts or eliminates the choice of individual.

Figure 3: House of Lords Table of Interventions Science et al. (2011, p.10)

Secondly, nudges themselves may act as a ban and nudge through regulatory means. Businesses can be required through regulation to present its services in a specific way, to nudge individuals. So a law can force a social actor to nudge its users, thereby counting both as a mandate and a nudge. The compatibility of nudging with the concerns of appropriateness might also have the positive benefit of highlighting the potential implications of the widespread use of nudges by commercial organizations, as was the case, mentioned in the introduction, with OECD's (International Congress on Transport and Infrastructure Systems) ban on pre ticked boxes.

6.2.2 Public Permission

(Dolan – Cabinet office) suggested that the ethical acceptability of intervention was related to its level of public acceptance or popularity. (reference). This link might seem unconvincing, since public acceptance might improve after the intervention has taken place, as with the general ban on indoor smoking or the Swedish alcohol monopoly “according to Gothenburg’s SOM Institute for public opinion research. Just 17 per cent favor lower taxes on alcohol, down from almost 60 per cent in 2005.” (Crouch, 2017). Also, this acceptance might be co-determined by the apparent success or failure of the intervention over time. It is thus entirely possible that government can implement ethically acceptable nudges even though they are unpopular if evidence can support it. At the same time, interventions which do not have public support

is, in general, less likely to succeed. The public perception of nudges is therefore relevant to investigate before enacting it as a policy decision. Nudges are often formulated as universally agreeable, the presentation of these initiatives often highlight how actors themselves agree to the desired outcome and the unproblematic implementation of nudge environment. Nudges in themselves are defined as being easy to resist/avoid, the logic and rationality of implementing nudges, on the other hand, might seem very hard to resist, especially taking into account the theory's apparent foothold as beneficial. There are however many aspects to consider when incentivizing individual behavior. One such aspect is the possible risk of reactance from people being nudged in a particular direction. Another aspect is the unknown interactions there is between hard power and soft power, and a related issue that soft power requires a certain level of hard power, a penalizing outcome (Science et al., 2011).

Along with several other scholars (Lodge & Wegrich, 2016) also question the longevity of nudge initiatives, this issue is currently mostly up to pure faith, despite nudging reputation for being evidence based, this apparent and immediate trustworthiness is probably derived from its attempts to combine micro economic beliefs about behavior with psychology. Any social situation, it is claimed, contains choice architectures and therefore nudges. Since we are always influenced by choice architecture, we should improve our decision quality. That is, by intelligently changing the choice architecture, humans are guided in their situation to act according to their preferences without suffering major opportunity costs (Sunstein & Thaler, 2008). Individuals continue to satisfice, but achieve more desirable outcomes, for themselves and the wider society. Also, they are usually not required to choose how to obtain the desired result. When a nudge has been designed, individuals are free to 'opt out' rather than 'opt in.' Thus, people eat healthier when given the opportunity to commit to a choice in a given future or encountering appealing fruit before chocolate bars and bags of crisps. Choice architecture suggests that those deciding on choice designs and dominant psychological mechanisms are equipped with perfect rationality, or, at least, superior rationality compared to the individual citizen (Rizzo, 2017).

6.2.3 Mandated choice or presumed consent

The forms of consent that are related to behavior change policies vary widely across the policies that we have encountered. Many nudge-type policies, particularly those related with the automatic enrolment of people onto organ donor registers and company pension schemes, adopt a presumed consent approach to behavior change. Presumed consent is based on the assumption that people want to donate their organs and join a company pension scheme, but for whatever reason simply never get around to doing so, to address this behavioral tendency, presumed consent policies simply reset the default position and place people on donor registers and pension schemes automatically.

Presumed consent policies, while often controversial, do not, of course, prevent people from choosing not to consent to the behavior in question, they only change the default position. At the other end of the spectrum, mandated choice policies seek to prompt actions not by resetting the default but by making it easier for people to make decisions. In the UK, for example, when applying for a new (or renewed) driving license, people are required to choose whether they would like to join the UK Organ Donor Register (Science et al., 2011). This tick box exercise is a useful psychological prompt to action but also makes it much easier to register for organ donation.

Behavioral economics has managed to highlight the behavioral alternative to information-based campaigns. The establishment of a variety of nudge institutions and units in Western nations reinforces the positive conception these ideas have brought with them. There is the creation of the Behavioural Insights Team (BIT) in 2010 in Great Britain, as well as the recruitment of behavioral scientists in British, American, French, Dutch and German ministries and agencies. A clear indication of the growing interest from public policymakers to broaden their toolkit (Oliver, 2013). Especially the BIT enjoyed a public praise due to its activities resulting in cost savings for the British government above £300 million compared to the £30 million budget of the team (Ly & Soman, 2013, p.13). Furthermore, a growing population, increasing societal costs for social and medical services and budgetary constraints raises significant interest in any technique that promises beneficial behavioral change without noteworthy costs. However, when nudging is implemented in a time of austerity, government experiments involving American-imported ideas, in this context, can become the focus of resentment and ridicule. For those provoked by funding cuts, these methods can be seen as part of an ideological plan to 'roll back' the state and make ordinary citizens and initiative pay for services themselves (Burgess, 2012, p.22).

6.2.4 Dubious Nudging

(Just & Wansink, 2014) found that individuals eat less when a food item is given a name that sounds big (like "*double-size*"). At the same time, individuals are willing to pay more for a good that sounds larger. One could imagine a policy in which larger names are given to food items without changing the choice set of available items. This policy may encourage individuals to eat less, though again it is difficult to use any standard version of welfare theory to demonstrate that the individual would be better off for this reduction. On the other hand, if such a policy encourages food producers to raise prices, it may improve the profits of producers at consumers' expense. Consumers may pay more for the same products and eat less of them. Under the constraint of finding a policy to encourage the consumption of less food, there may be some way to argue a welfare improvement over more directly paternalistic policies. However,

without such a constraint, it is difficult to see how a clear argument could be made to justify such a policy using traditional welfare theory.

6.2.5 Political Critique

When used by policymakers, nudges usually are targeted on groups of people, namely a certain population. Thaler and Sunstein, however, do not state any further political goals that government may pursue by applying nudging other than enabling “*better governance*” [93]. This better governance is again characterized as less “*in the way of government coercion and constraint*” (Sunstein & Thaler, 2008, p.14) and both smaller and more modest than previous models of government. Thus, they are only referring to the way power is exercised by the government and not to ulterior goals. However, through the academic replies to nudge theory, it became evident that the technique of nudging is accused of promoting a distinct political agenda. By endorsing maximization of rationality within individual choices as an objective with intrinsic value, nudging no longer qualifies as a neutral, governance-enhancing tool.

As a pre-given and objective rationality of the individual is negated within Foucauldian theory, it has to be clarified which subjective rationality is promoted by nudging. It is not the subjective and individual rationality of the citizen, but the governmental reason of the state that should be promoted to be decisive in citizens’ situations of choice. It is clear that nudges are in line with a political rationality driven by the logic of political economy, aiming at the improvement of health, wealth and the subsistence of its population, is promoted. This statement rests on several chains of thought. First, nudges such as the setting of defaults do convey a certain normative message. Within their examples, people usually are nudged towards a more healthy, environmentally-friendly or financially sustainable behavior.

Even though it is argued by (Sunstein & Thaler, 2008), that such nudges mainly are beneficial for the individual being nudged, it has to be considered that there are nudges that do not bear any direct benefit for the individual. A nudge like *Don’t mess with Texas!* Sunstein and Thaler (2008, p.60), consisted in broadcasting a slogan addressing the local pride of citizens to reduce littering. This initiative is mainly driven by concerns for the population’s welfare, rather than individual benefits. Rather, via nudging ulterior government motives aiming at the whole nation or population were pursued.

6.2.6 The Current Status of Nudging

After reviewing the most popular contributions to the debate on nudging it becomes apparent that there is no broad consensus within the scientific community regarding the interpretation of the exercise of power of nudges. Libertarian critique interprets nudging mainly as a disturbance

of originally independent decision-making processes and therefore as a violation of individual liberty. This group has a highly critical stance towards the idea of man proposed by Thaler and Sunstein and instead highlights the difficulty of exogenously defining and measuring rationality.

Liberty, often defined by the absence of foreign influences, seems to be endangered by nudges if they do not fulfill strict conditions stated by the theoreticians. Libertarian critics share a similar language, as they warn of the manipulative and paternalistic traits that nudges may bear. Anti-libertarian and deliberative replies as well do not exuberantly support the application of nudging techniques but have a more positive stance. They base their criticism on a positive definition of liberty, highlighting the necessity of nudges being empowering and leading to autonomous deliberation. Especially deliberative authors support the application of nudges, given they include an element of empowerment. To conclude, the prominent replies to Thaler and Sunstein's publication do primarily discourage the extension of nudging techniques due to concerns about violations of liberty and autonomy. This may in consequence exercise an adverse influence on the application of nudging techniques within the sphere of policy-makers. It has to be admitted though that both the spread of nudging institutions as well as Foucauldian theory indicate the opposite: on the one hand, governmental organizations engaged in nudging are established on a regular basis, as depicted in part 2. On the other hand, the analysis of the power of nudges revealed an enormous potential for government intervention. Nudging techniques will not replace legalistic measures of discipline. However they bear a higher potential of influencing citizens, as they proceed subliminally and therefore are harder to avoid.

6.3 Conclusion

This thesis has taken a step back from the heated normative discussions about nudges and focused on conceptual and methodological issues surrounding nudges instead. This way of describing nudges has allowed for a discussion of various aspects of choice architecture. The thesis has asked what kinds of assumptions are needed to endorse nudges completely, both to normatively endorse the assumptions of successful nudges, as well as the additional conditions for the implementation, transparency, and justification of nudges. The presentation put forward in this thesis makes inroads with regards to where exactly the methodological and normative problems with nudges lie. Either in the specific assumptions necessary for successful nudges, or in the general conditions and context of their implementation.

For instance, one might be concerned with privacy when a choice architect (such as a government agency or contractor) is making assumptions about the mental state of decision-makers on the level of detail that is required for nudges. Making assumptions of this kind might also prompt the choice architect to learn more about these types of dispositions in the first place, and while governments and their subsidiaries are trying to learn a great deal about citizens. Choice

architects should pay attention to the vast type of nudging interventions available to them and consider the transparency of interventions with their potential to infringe on privacy.

Moreover, even if one is prepared to endorse nudges in principle, one might judge it as likely to turn into manipulation or believe that it places unreasonable requirements to a choice architect capabilities and therefore not possible to endorse in practice. Nudge that relies on sound empirical research to establish correct assumptions about the initial state of mind, and which also makes it easy for individuals to which the assumption does not hold to ‘oppose’ the Nudge, may well have less interference with the autonomy of decision-makers than social advertising.) Such considerations will not make the normative problems of nudging disappear, but they offer a way to clarify what kinds of problems with nudges could be rectified by making sure to provide transparent nudges. In (Hausman & Welch, 2010) ’s case, it might be the case that they would still reject nudges even if it were successful in the sense defined in chapter 4, because they see it as a type of paternalism that is not defensible. However, their position might also partly be due them finding it implausible that nudges can be sufficiently transparent.

The second rejection of nudging deals with the fact that nudge interventions can quickly turn into manipulations and other types of interventions that cannot be endorsed. As examples of this, take (Hausman & Welch, 2010, p.131) as well as (Grüne-Yanoff, 2012, p.636) criticisms of nudges because they can easily become manipulative. This thesis has engaged in this type of problems by explicitly presenting the assumption behind nudges, if the nudge results in a choice that the decision-maker would have reflectively endorsed, then the nudge is used in contradiction with its original purpose. In chapter 4, the thesis noted that this assumption is adamant, and the normative critique of the authors mentioned above seems to hang on this. The upshot of this discussion is that there are some strong assumptions amongst those that endorse nudges, and that the strength of these assumptions makes it normatively problematic to endorse nudges (Heilmann, 2014) . Thaler and Sunstein repeatedly argue for welfare-promoting nudges by saying that the individuals are already influenced by choice architecture in their everyday choices. This inevitability argument has attracted much criticism in the normative discussion. As the objection goes, it is not the same, to be influenced by an element in the choice conditions without anyone, in particular, being in control as for when someone uses this element for a purpose (Grüne-Yanoff, 2012; Hausman & Welch, 2010) . Without dwelling further into this important debate, one should note that using the same word for the intervention and its underlying mechanism can only foster confusion between the two cases, and this is a good reason for avoiding this language.

Some authors contend that there are some normative problems which are not associated with nudges working correctly, but rather with what kind of impact they may have, and how they relate to or contradict other important goals that policymakers may want to promote. These pertain to problems of implementation, transparency, and justification, which have received the

lion's share of coverage in the normative debates about nudges. For instance, (Bovens, 2009, pp.211-14) speculates that nudges can infringe on the autonomy of decision-makers in the long-run, as they may erode the capacities of reflection and using their deliberative faculties. Another worry about nudge is that it can lead to abuse by authorities, such as social planners and governments (Hausman & Welch, 2010, p.136). . In a similar vein, some commentators worry that nudges are difficult to implement such that transparency and publicity can be guaranteed (Bovens, 2009, p.216) and (Hausman & Welch, 2010, p.135) . Finally, some observers worry that nudges do not have libertarian or liberal credentials: (Hausman & Welch, 2010, p.136) raise doubts about the libertarian credentials of nudges and (Grüne-Yanoff, 2012) believes that nudges violate core liberal principles. It seems that this type of criticism raises broader concerns about the impact, institutional conditions of implementation, transparency and general justification of nudges.

Some methodological principles for the design of nudges could be derived to mitigate some of the criticisms. For instance, a strong requirement for inducing a genuine nudge position that is, a position in which an individual can 'opt-out' of or 'oppose' the nudge. Respecting these requirements in the implementation of nudges might persuade some critics who worry about the erosion of capacities of decision-makers and those who worry about transparency. Indeed, while this thesis has tried to offer a clarification in a constructive spirit, it seems also to have the normative ideal of decision-making is indistinguishable from rational choice models. The criticism of nudging implied in this way seems at least as important as the 'purely' normative ones that begin by considering concepts such as libertarianism, freedom of choice, paternalism, manipulations and asks whether they apply to nudges. That is to say, this thesis has attempted to stay as close as possible to what one can reasonably assume is the core idea of nudges when criticising the idea and found that problems concerning manipulation, freedom of choice and erosion of agency can partly be seen as problems stemming from the way nudges are designed.

The finding does not preclude a valid normative criticism of the more general kind. Rather, it points to ways in which one can discuss normative problems of nudges in close relation to what they purport to do. The picture developed here also suggests that one should ask when, why and why issues arise and whether they can be resolved. Rather than simply asking whether nudges are right or wrong, in the abstract. It seems clear that the normative problems related to nudging cannot be divided into 'transparent' on the one hand and 'non-transparent' on the contrary. There are still two different questions for further research. The first is asks under what conditions do nudges work like they are supposed to work, and the second asks whether in each particular case can nudges be endorsed normatively. Making inroads on the first question has been the main goal of this thesis. The latter seems equally important as it highlights the distinctive character of nudging interventions. It lies in the following: even if there is still considerable work to do, the account in this thesis suggests that nudge policies are interventions that are com-

paratively easy to analyze with regards to the assumptions they make about the individuals in society. All this does not necessarily make nudges better than other types of policies: many have argued that it makes matters worse. On a descriptive and conceptual level, nudging allows for an increasingly learned disagreement over details than conventional policies. In this sense, it seems that more traditional ways of policy-making operate in a much less transparent way than nudges concerning their underlying conceptual and methodological assumptions. In this thesis, the intellectual origins and implications of nudging have been reviewed. The goal of this thesis has been to convey a major distinction between ways of thinking about behavior change, where nudging is concerned with providing cues to individuals to improve society. Thaler and Sunstein acknowledges the limits of individual capabilities and the lack of capacity individuals have in their everyday lives and decision-making for weighing up all the options that are open to them, subsumed in bounded rationality.

The arguments developed in this thesis are not objections to a version of behavioral welfare economics that claims to regularise revealed preferences inconsistent with conventional rational agent theory. However, this is not the version of nudging that is to be found in the literature. In arguing for nudging Thaler and Sunstein criticise conventional economists understanding individuals as 'Econs.' Thaler and Sunstein claim that their approach, libertarian paternalism, to behavioral welfare economics breaks apart from the unrealistic assumption, and models human decision-making as it is. This thesis has argued that Thaler and Sunstein model people as incomplete 'Econs'. Their normative model of human decision-making is rational choice theory combined with a context-dependent individual. Their goal of nudging interventions is primarily an attempt to reconstruct and respect the preferences of the imagined 'Econ'. This thesis has argued that Thaler and Sunstein model people as incomplete 'Econs.' Their normative model of human decision-making is rational choice theory combined with a context-dependent individual. Their goal of nudging interventions is primarily an attempt to reconstruct and respect the preferences of the imagined 'Econ.' The thesis maintains that if behavioral and normative economics is to be satisfactorily reconciled. Thaler and Sunstein should dismiss their "*as judged by themselves*" criteria, which attempts to reconstruct the latent neoclassical preferences of individuals whose psychology causes mistakes. Thaler and Sunstein might think of themselves as trying to represent the complex reality of human judgment and decision-making in a highly simplified normative modeling framework. However, that would require a significant retreat from the ambition of much of the current literature of behavioral welfare economics.

Nudges will increasingly be applied in public policy, but one has to be careful with its application and how it might be perceived, in this regard transparency is important to avoid accusations of manipulation or suspecting nudges for simple austerity measures. Research is still in the early stages and lacks knowledge when it comes to the concrete application and the long-term effects of various nudging initiatives. One must also take into account the likelihood that a

given nudge will create reactance from the people being nudged, and evaluate whether a certain outcome is desirable, from an individual standpoint and a policy point. To understand nudges one must also understand its ontological and epistemological foundations. This will illuminate the strength and weaknesses in nudging and help to guide nudging into a technical application, instead of being kidnapped by politically oriented ideas. Nudges are placed on a continuum between positive and normative aspirations and should ideally require public debate. Debating the virtues of nudging is the most appropriate way to maximize the principles presented by Thaler and Sunstein in evaluating nudges.

7 Bibliography

References

- Adams, T. (2014, 06). Nudge economics: has push come to shove for a fashionable theory? *theguardian*. Retrieved from <https://www.theguardian.com/science/2014/jun/01/nudge-economics-freakonomics-daniel>
- Akerlof, G. A. (1991). Procrastination and obedience. *The American Economic Review*, 81(2), 1–19.
- Akerlof, G. A., & Shiller, R. J. (2015). *Phishing for phools: The economics of manipulation and deception*. Princeton University Press.
- Akrich, M., & Latour, B. (1992). *A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies*. The MIT Press.
- Amir, O., & Lobel, O. (2008). Stumble, predict, nudge: How behavioral economics informs law and policy. *Columbia Law Review*, 2098–2137.
- Backhouse, R. E., & Backhouse, R. (2002). *The penguin history of economics*. Penguin UK.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American psychologist*, 54(7), 462.
- Beck, A. T. (1964). Thinking and depression: Ii. theory and therapy. *Archives of General Psychiatry*, 10(6), 561-571. Retrieved from <http://dx.doi.org/10.1001/archpsyc.1964.01720240015003> doi: 10.1001/archpsyc.1964.01720240015003
- Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond*. Guilford Press.
- Bentham, J. (1791). *Panopticon or the inspection house* (Vol. 2; M. Bozovic, Ed.). Transcription and HTML by Cartome.
- Bergstrom, T. C. (2002). Evolution of social behavior: individual and group selection. *The Journal of Economic Perspectives*, 16(2), 67–88.
- Blaug, M. (2017, 04). economics. *Encyclopaedia Britannica*. Retrieved from <https://www.britannica.com/topic/economics/Methodological-considerations-in-contemp>
- Blumenthal-Barby, J. S., & Burroughs, H. (2012). Seeking better health care outcomes: the ethics of using the nudge. *The American Journal of Bioethics*, 12(2), 1–10.
- Bovens, L. (2009). The ethics of nudge. In *Preference change* (pp. 207–219). Springer.
- Burgess, A. (2012). Nudging healthy lifestyles: The uk experiments with the behavioral alternative to regulation and the market. *Eur. J. Risk Reg.*, 3, 3.
- Burns, Z., Chiu, A., & Wu, G. (2010). Overweighting of small probabilities. *Wiley Encyclope-*

dia of Operations Research and Management Science.

- Buyx, A. M. (2009). Blood donation, payment, and non-cash incentives: classical questions drawing renewed interest. *Transfusion Medicine and Hemotherapy*, 36(5), 329–339.
- Byrne, M. D., & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive science*, 21(1), 31–61.
- Casey, B., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., ... others (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences*, 108(36), 14998–15003.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.
- Cohen, S. (2013). Nudging and informed consent. *The American Journal of Bioethics*, 13(6), 3–11.
- Crawford, S. E. S., & Ostrom, E. (1995). A grammar of institutions. *The American Political Science Review*, 89(3), 582-600. Retrieved from <http://www.jstor.org/stable/2082975>
- Crouch, D. (2017). *Sweden's strict controls drive 'booze-cruise' economy*. Retrieved from <https://www.ft.com/content/6040e408-ed33-11e6-ba01-119a44939bb6>
- Daniel Kahneman, J. L. K. (2000). Experimental tests of the endowment effect and the coase theorem(second order decisions). In C. R. Sunstein (Ed.), *Behavioral law and economics* (p. 211-232). Cambridge University Press.
- Dobson, A. (2007). Environmental citizenship: towards sustainable development. *Sustainable development*, 15(5), 276–285.
- Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, 33(1), 264 - 277. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0167487011001668>
doi: <http://dx.doi.org/10.1016/j.joep.2011.10.009>
- Dubov, A. (2015). Ethical persuasion: the rhetoric of communication in critical care. *Journal of evaluation in clinical practice*, 21(3), 496–502.
- Dutta, P. K. (1999). *Strategies and games: theory and practice*. MIT press.
- Egan, T. (2005). Newly bankrupt raking in piles of credit offers. *New York Times*, 1.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, 59, 255–278.
- Fateh-Moghadam, B., & Gutmann, T. (2014). Governing [through] autonomy. the moral and legal limits of “soft paternalism”. *Ethical Theory and Moral Practice*, 17(3), 383–397. Retrieved from <http://dx.doi.org/10.1007/s10677-013-9450-3> doi: 10.1007/s10677-013-9450-3
- Fong, C. M., Bowles, S., & Gintis, H. (2006). Strong reciprocity and the welfare state. *Hand-*

- book of the economics of giving, altruism and reciprocity*, 2, 1439–1464.
- Foucault, M. (1982). The subject and power. *Critical inquiry*, 8(4), 777–795.
- Foucault, M. (1990). The history of sexuality: An introduction, volume i. *Trans. Robert Hurley*. New York: Vintage.
- Foucault, M. (1991). *The foucault effect: Studies in governmentality* (C. Gordon, G. Burchell, & P. Miller, Eds.). Wheatsheaf.
- Foucault, M. (2007). *Security, territory, population: lectures at the collège de france, 1977-78*. Springer.
- Foucault, M. (2012). *Discipline & punish: The birth of the prison*. Vintage.
- Francesco, G. (2010). Cooperation in and out of the lab: a comment on binmore's paper. *Mind & Society*, 9(2), 159-169.
- Freidman, M. (1994). *Readings in the philosophy of social science chapter 41: The methodology of positive economics* (M. Martin & L. C. McIntyre, Eds.). Mit Press.
- French, J. (2011). Why nudging is not enough. *Journal of Social Marketing*, 1(2), 154-162. Retrieved from <https://doi.org/10.1108/20426761111141896> doi: 10.1108/20426761111141896
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62, 451–482.
- Gosseries, A., & Parr, T. (2017). *The stanford encyclopedia of philosophy: Publicity* (Fall 2017 ed.). Metaphysics Research Lab, Stanford University. Retrieved from <https://plato.stanford.edu/archives/fall2017/entries/publicity/>
- Goyle, A. G. (2017). *prison*. Website. Retrieved from <https://www.britannica.com/topic/prison>
- Grüne-Yanoff, T. (2012). Old wine in new casks: libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, 38(4), 635–645.
- Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *Eur. J. Risk Reg.*, 3.
- Harris, R., & Albin, E. (2006). Bankruptcy policy in light of manipulation in credit advertising. *Theoretical inquiries in Law*, 7(2), 431–466.
- Hausman, D. M. (2011). *Preference, value, choice, and welfare*. Cambridge University Press. doi: 10.1017/CBO9781139058537
- Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy*, 18(1), 123–136.
- Heilmann, C. (2014). Success conditions for nudges: a methodological critique of libertarian paternalism. *European Journal for Philosophy of Science*, 4(1), 75–94. Retrieved from <http://dx.doi.org/10.1007/s13194-013-0076-z> doi: 10.1007/s13194-013-0076-

Hicks, J. R. (1939). The foundations of welfare economics. *The Economic Journal*, 49(196), 696–712.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science*, 4(2), 162–176.

Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Lewis, H. B., Wei, Y., ... Ogilvie, D. (2015). Portion, package or tableware size for changing selection and consumption of food, alcohol and tobacco. *The Cochrane Library*.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, 9(1), 90–98.

Ilana Ritov, J. B. (2000). *Behavioral law and economics: Chapter 6 reluctance to vaccinate: Omission bias and ambiguity* (C. R. Sunstein, Ed.). Cambridge University Press.

James, O., & Moseley, A. (2014). Does performance information about public services affect citizens' perceptions, satisfaction, and voice behaviour? field experiments with absolute and relative performance information. *Public Administration*, 92(2), 493–511.

John, P., Cotterill, S., Moseley, A., Richardson, L., Smith, G., Stoker, G., & Wales, C. (2011). *Nudge, nudge, think, think: Experimenting with ways to change civic behaviour* (1st ed.). London: Bloomsbury Academic. Retrieved from <http://www.bloomsburycollections.com/book/nudge-nudge-think-think-experimenting-wit>

Jones-Lee, M., & Loomes, G. (2001). Private values and public policy. *Conflict and tradeoffs in decision making*, 205–230.

Jung, J. Y., & Data, B. A. M. (2016). American attitudes toward nudges. *Judgment and Decision making*, 11(1), 62.

Just, D. R. (2017). The behavioral welfare paradox: Practical, ethical and welfare implications of nudging. *Agricultural and Resource Economics Review*, 46(1), 1–20.

Just, D. R., & Wansink, B. (2014). One man's tall is another man's small: how the framing of portion size influences food choice. *Health economics*, 23(7), 776–791.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American economic review*, 93(5), 1449–1475.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives*, 5(1), 193–206.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, 263–291.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American psychologist*,

39(4), 341.

- Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal*, 549–552.
- Klein, G. A. (1999). *Sources of power: How people make decisions*. MIT press.
- Koeszegi, B., & Rabin, M. (2007). Mistakes in choice-based welfare analysis. *The American Economic Review*, 97(2), 477–481. Retrieved from <http://www.jstor.org/stable/30034498>
- Levin, J., & Milgrom, P. (2004). Consumer theory. Retrieved May, 24, 2015.
- Lodge, M., & Wegrich, K. (2016). The rationality paradox of nudge: Rational tools of government in a world of bounded rationality. *Law & Policy*, 38(3), 250–267.
- Loersch, C., & Payne, B. K. (2011). The situated inference model: An integrative account of the effects of primes on perception, behavior, and motivation. *Perspectives on Psychological Science*, 6(3), 234–252.
- Loewenstein, G. (1999). Because it is there: The challenge of mountaineering for utility theory. *Kyklos*, 52(3), 315–343.
- Ly, K., & Soman, D. (2013). Nudging around the world. *Behavioral Economics in Action@Rotman School of Management*.
- Marchiori, D. R., Adriaanse, M. A., & De Ridder, D. T. (2017). Unresolved questions in nudging research: Putting the psychology back in nudging. *Social and Personality Psychology Compass*, 11(1).
- Mc Culloch, K. C., Aarts, H., Fujita, K., & Bargh, J. A. (2008). Inhibition in goal systems: A retrieval-induced forgetting account. *Journal of experimental social psychology*, 44(3), 857–865.
- Mik-Meyer, N., & Villadsen, K. (2007). *Magtens former*. Hans Reitzel.
- Mills, C. (2013). Why nudges matter: A reply to goodwin. *Politics*, 33(1), 28–36.
- Mueller, D. C. (2004). Models of man: neoclassical, behavioural, and evolutionary. *politics, philosophy & economics*, 3(1), 59–76.
- Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The pull of the past: When do habits persist despite conflict with motives? *Personality and Social Psychology Bulletin*, 37(11), 1428–1437.
- OECD. (2017, 05). Why bother? context and approach. *OECD Publishing*, 15-30. Retrieved from /content/chapter/9789264270480-4-en doi: <http://dx.doi.org/10.1787/9789264270480-4-en>
- Oliver, A. (2011). A nudge too far? a nudge at all? on paying people to be healthy. *Health-carePapers*, 12(4), 8–16.
- Oliver, A. (2013). From nudging to budging: using behavioural economics to inform public sector policy. *Journal of Social Policy*, 42(04), 685–700.

- Pacuit, E., & Roy, O. (2017). Epistemic foundations of game theory. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. Retrieved from <https://plato.stanford.edu/archives/sum2017/entries/epistemic-game/>
- Quinn, J. M., Pascoe, A., Wood, W., & Neal, D. T. (2010). Can't control yourself? monitor those bad habits. *Personality and Social Psychology Bulletin*, 36(4), 499–511.
- Rebonato, R. (2014). A critical assessment of libertarian paternalism. *Journal of Consumer Policy*, 37(3), 357–396. Retrieved from <http://dx.doi.org/10.1007/s10603-014-9265-1> doi: 10.1007/s10603-014-9265-1
- Rizzo, M. J. (2017). Rationality—what? misconceptions of neoclassical and behavioral economics. *Behavioural Public Policy, Forthcoming; HEC Paris Research Paper*.
- Rostain, T. (2000). Educating homo economicus: cautionary notes on the new behavioral law and economics movement. *Law and Society Review*, 973–1006.
- Rumelt, R. P. (2012). Good strategy/bad strategy: The difference and why it matters. *Strategic Direction*, 28(8).
- Russo, J. E., Schoemaker, P. J., & Russo, E. J. (1989). *Decision traps: Ten barriers to brilliant decision-making and how to overcome them*. Doubleday New York.
- Saghai, Y. (2013). Salvaging the concept of nudge. *Journal of medical ethics*, 39(8), 487–493.
- Samson, A. (2014). *The behavioral economics guide 2014*. Behavioral Science Solutions.
- SBST. (2016). Social and behavioral science team annual report. *Executive office of the President, National Science and Technology Council, Washington, D.C.*
- Science, Committee, T., et al. (2011). House of lords-science and technology select committee 2nd report of session 2010-12—behaviour change. *House of Lords Paper*, 179.
- Sen, A. (1973). Behaviour and the concept of preference. *Economica*, 40(159), 241-259. Retrieved from <http://www.jstor.org/stable/2552796>
- Sen, A. K. (2014). *Collective choice and social welfare* (Vol. 11). Elsevier.
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99-118. Retrieved from <http://www.jstor.org/stable/1884852>
- Smith, M. A., & McPherson, M. S. (2009). Nudging for equality: Values in libertarian paternalism. *Administrative Law Review*, 61(2), 323-342. Retrieved from <http://www.jstor.org/stable/41553553>
- Spence, C., Harrar, V., & Piqueras-Fiszman, B. (2012). Assessing the impact of the tableware and other contextual variables on multisensory flavour perception. *Flavour*, 1(1), 7.
- Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature*, 38(2), 332–382.
- Sugden, R. (2009). *On nudging: A review of nudge: Improving decisions about health, wealth and happiness by richard h. thaler and cass r. sunstein*. Taylor & Francis.

- Sunstein, & Thaler. (2008). *Nudge: Improveing decisions about health, wealth, and happiness*. Yale University press.
- Sunstein, C., R. (2017). *Don't tell me what i can't do!' on the intrinsic value of control*. Retrieved from https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2896949 (Introduction, Behavioral Economics Guide)
- Sunstein, C. R. (2000). *Behavioral law and economics: Chapter 7 second order decisions*. Cambridge University Press.
- Sunstein, C. R. (2014). *Why nudge?: The politics of libertarian paternalism*. Yale University Press.
- Sunstein, C. R. (2016a). Do people like nudges? *Administrative Law Review*.
- Sunstein, C. R. (2016b). *The ethics of influence: Government in the age of behavioral science*. Cambridge University Press.
- Sunstein, C. R. (2017). *# republic: Divided democracy in the age of social media*. Princeton University Press.
- Sunstein, C. R., & Reisch, L. A. (2014). Automatically green: Behavioral economics and environmental protection. *Harv. Envtl. L. Rev.*, 38, 127.
- Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 1159–1202.
- Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making*, 12(3), 183.
- Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of political Economy*, 112(S1), S164–S187.
- Vallentyne, P., & van der Vossen, B. (2014). *Libertarianism*. The Stanford Encyclopedia of Philosophy. Retrieved from <https://plato.stanford.edu/entries/libertarianism/>
- Wansink, B., & Hanks, A. S. (2013). Slim by design: serving healthy foods first in buffet lines improves overall meal selection. *PloS one*, 8(10), e77055.
- Wansink, B., Painter, J. E., & North, J. (2005). Bottomless bowls: why visual cues of portion size may influence intake. *Obesity*, 13(1), 93–100.
- Wansink, B., & Van Ittersum, K. (2013). Portion size me: Plate-size induced consumption norms and win-win solutions for reducing food intake and waste. *Journal of Experimental Psychology: Applied*, 19(4), 320.
- Whitehead, M., Jones, R., Howell, R., Lilley, R., & Pykett, J. (2014). Nudging all over the world. *ESRC Report, Economic and Social Research Council, Swindon and Edinburgh*.
- Wilkinson, T. M. (2013). Nudging and manipulation. *Political Studies*, 61(2), 341–355.
- Wood, W., Labrecque, J. S., Lin, P.-Y., & Runger, D. (2014). Habits in dual process models. *Dual process theories of the social mind*, 371–385.