When are nudges acceptable?

Influences of beneficiaries, techniques, alternatives and choice architects

William Hagman

LI.U

LINKÖPING UNIVERSITY

At the Faculty of Arts and Sciences at Linköping University, research and doctoral studies are carried out within broad problem areas. Research is organized in interdisciplinary research environments and doctoral studies mainly in graduate schools. Jointly, they publish the series Linköping Studies in arts and Science. This thesis comes from the Division of Psychology at the Department of Behavioural Sciences and Learning.

William Hagman
When are nudges acceptable?
Influences of beneficiaries, techniques, alternatives and choice architects

To my wife, the great choice architect in my life.

Contents

Acknowledgment

I would not have been able to finish my PhD studies without the help and support from my family, friends and colleagues and I want to thank you deeply. I would like to give a special thanks to my supervisor, Professor Daniel Västfjäll, for always letting me follow my interests and encourage my work. I would also like to thank my co-supervisor Stephan Dickert for good discussions and valuable feedback.

To all members of the JEDI-lab, I am grateful. All our fun, after work activities have also given me friends for life. The conversations, seminars, and workshops have improved my projects and a special thanks to Lina Koppel for proofreading the draft of the thesis.

Additionally, I would like to thank my colleagues both at IBL and at the Economics Division at Linköping University for all the support you have given me.

Abstract

Interventions aimed to change behavior (so called nudges) are becoming more and more popular among policymakers. However, in order to be able to effectively use nudges, it is important to understand when and why people find them acceptable. The objective of this thesis is therefore to improve the understanding of when nudges are judged to be acceptable. The thesis focuses on a model for behavioral change. The model contains two parts, nudge technique and acceptance of nudges. Nudge technique refers to how the nudge is designed to function in regard to psychological mechanism and functionality.

The nudge technique part of the model is expanded and problematized from an ethical perspective in the first part of this thesis, by exemplifying psychological mechanisms behind different techniques and explaining why they might be intrusive to individuals' freedom of choice. In the second part of this thesis it is discussed why acceptance is an important component of making nudging legitimate and effective. This is followed by a discussion of how acceptance is empirically measured. The empirical part of the thesis is based on four papers which all use a quantitative online survey approach to study the judgements of nudges from the general public.

Paper 1 was a first attempt to measure whether nudges which are common in the nudge literature are acceptable interventions according to the general public. We found that the nudges that were categorized as pro-self were more likely to be rated as acceptable and less likely to be perceived as intrusive to freedom of choice compared to pro-social nudges. Furthermore, the effect of decision styles and worldview on acceptance was explored. In paper 2, we explored whether the difference between acceptance found for pro-social nudges and pro-self nudges could be increased by framing nudges as beneficial for society or individuals. The framing had no effect on acceptance but, as in paper 1, pro-social nudges were found to be more intrusive to freedom of choice compared to pro-self framed nudges. Moreover, different nudge techniques had different rates of acceptance even with the same explicit goal for the nudges. In paper 3, we examined whether the alternative to nudges affects the perceived acceptability and intrusiveness of default-changing nudge techniques. The alternatives given to the nudges were either to enforce the intended behavioral

change with legislation or to do nothing at all in order to change the behavior. We find no difference in aggregated acceptance, however, the judgements vary depending on individuals' worldview. Paper 4 explored if the choice architect's (the creator/proposer of the nudge) political affiliation affects acceptance rating for proposed nudge interventions and legislation. We find that acceptance of both nudges and legislation increases with the level of matching between people's political orientation and the choice architect's political affiliation.

Taken together, the findings suggest that there is more to creating an acceptable nudge than to merely take a nudge technique that was acceptable in one context and apply it in another. Moreover, nudges that are rated as more beneficial towards individuals compared to society at large are in general more likely to be found acceptable and less intrusive to freedom of choice. It is important to have knowledge about the target population (e.g. their decision styles, world-views, and political orientation) to avoid backfires when implementing nudges.

List of papers

1: Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public views on policies involving nudges. *Review of Philosophy and Psychology*, *6*(3), 439-453.

2: Hagman, W., Erlandsson, A., Tinghög, G., & Västfjäll, D. Acceptability of nudges: The role of nudge technique and intended beneficiary. Under review.

3: Hagman, W., Tinghög, G., Erlandsson, A., Dickert, S., & Västfjäll, D. The effect of paternalistic alternatives on attitudes toward nudges. Under review.

4: Hagman, W., Tinghög, G., Erlandsson, A., Dickert, S., & Västfjäll, D. Partisan bias for acceptability of nudges in Sweden: When partisan-consistent legislation is as good as partisan-inconsistent nudges. Manuscript.

Introduction

Nudging is becoming a common tool for policymakers to use in order to steer individuals' behavior. A nudge is when knowledge of psychological mechanisms is being used to structure a choice setting. In this thesis I first present a very simple model of nudges and when they can change behavior (see Figure 1). This model consists of three concepts, the first is nudge technique which describes how the nudge is designed and used, the second is acceptance which describes factors that affect the acceptability of a nudge (which is empirically measured in this thesis), and the third is the outcome of the nudge, i.e. the behavior changes. The focus of this work is on the second concept and expanding the parts within it. However, in order to discuss acceptance of nudges is it important to describe the first part of the model, which is the design of the actual nudge. The first concept (nudge techniques) of the model is divided into smaller components which are problematize from an ethical perspective, before moving on to acceptance. Many theoretical concerns have been raised regarding nudges, for instance if nudge is manipulative by exploiting psychological mechanism to affect individuals' choices, and therefore intrusive to freedom of choice. The starting point of this thesis was to empirically test whether some of the theoretical concerns regarding nudges could make them unacceptable by the general public, and also to expand the second part of the model with factors that do affect the acceptability of nudges. Different designs of nudges with different goals are empirically tested to see if and how this affect acceptance of nudges.

Public acceptance of nudges can be measured in many different ways. For instance, it could be argued that acceptance should be measured by compliance to the intended behavior change. However, this is problematic for two reasons. The first is that poor design of nudges might be mistaken for nonacceptance. The other reason is that some nudge techniques might go unnoticed by the individual being nudged when the nudging occurs, which in turn could mistake obliviousness with acceptance. Moreover, since some individuals might be oblivious to the nudging you could argue that acceptance is not necessarily needed at all for nudges. However, when people find out that their behavior has been intentionally tampered with "for their

own good" it would not be far-fetched to assume that they will find this to be manipulative, paternalistic and probably not acceptable. Although acceptance of nudges could be seen as a proxy for behavior change is it not safe to assume acceptance of nudges only based on behavior. The need for acceptance of nudges are important for two reasons, without it will intended behavioral change be less likely to succeed and secondly, it is hard to argue that a society that use unacceptable nudges (i.e. manipulation) to change the behavior of the citizens are to be considered a democratic society. Another way to measure acceptance might be with a more qualitative approach, using interviews, observations (Lapan et al., 2001 and other design evaluation methods for instance cognitive walkthrough (e.g. Wharton et al, 1994). While this is a good approach to get an in-dept knowledge why a few people find nudges acceptable it is hard to get an appropriate sample size to draw general conclusions regarding the acceptance of nudges from the general public. Therefore, a web-survey method with a questionnaire has been used to measure whether people found the nudges to be acceptable and intrusive to freedom of choice in this thesis.
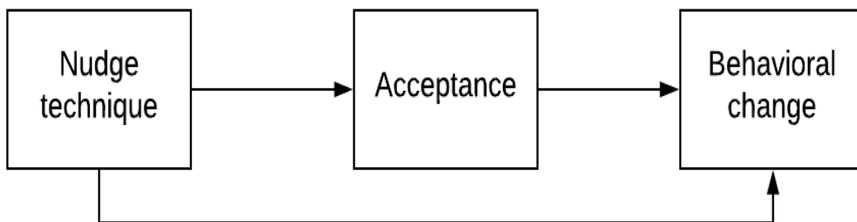
Figure 1. Model of behavioral change with nudges

In this simple model behavioral change through nudges have two necessary components: 1) a well-designed nudge technique and 2) acceptance of that nudge. This model is expanded throughout the thesis.

The outline of the thesis

The present thesis is divided into three parts, nudge technique, acceptance and general discussion. The thesis begins with the first part

explaining what the idea behind nudges is and why people would need them. Thereafter the psychological mechanisms and cognitive perspectives are presented to give an idea of how nudges can be used along with the principles of nudges. Then different components of nudge techniques are presented and discussed. Nudges are problematized from the perspective of paternalism, acceptance and individuals' preferences, and a reason why nudges might fail is presented. The second part of the thesis focuses on factors that affect acceptance of nudges. The four empirical papers are summarized one by one and the novel findings are included in a model for acceptance of nudges. The second part ends with presenting the final model of nudge techniques and acceptance. In the third part the contributions of the papers to the nudge literature are discussed, followed by a discussion regarding the future of nudges. The thesis is finished with concluding remarks and the full version of each paper.

What is the idea behind nudging?

Nudging has been widely acknowledged and "nudge-units" that work with policy recommendations have been emerging all over the world (e.g. in the US[1], UK[2] and the EU[3]). The term nudge has been wildly spread (and debated) both in academic settings but also in policy making. When nudges were introduced by Thaler and Sunstein in 2008, a nudge was defined as:

> … 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 fruit at eye level counts as a nudge. Banning junk food does not. – Thaler and Sunstein (2008) p 6.

The idea of nudges is to take concepts that behavioral economists, cognitive ergonomists and psychologists have researched for years and applying them in the policy settings in order to improve individuals' decision making. The reason that nudges are gaining fame arguably has two aspects. The first aspect is that nudging is an addition to other policy tools (e.g. bans, regulations or subsidies) to help individuals to stay away from undesired behaviors. Nudges are cheaper to test and implement compared to other regulations and are also easier to remove if the desired results are not achieved.

The second aspect, which can be considered the key component of nudges as presented by Thaler and Sunstein, is that a nudge is supposed to help the person being nudged to engage in a behavior that is in accordance with the person's own underlying preferences. In other words, the person who is being nudged is the person who should benefit from the nudge in such a way that is deemed beneficial for the individual by the individual himself (Sunstein, 2014; Sunstein, 2016a). This is an important feature of nudges that separates nudges from

---

[1] Social and Behavioral Sciences Team 2016 Annual Report; https://sbst.gov
[2] The Behavioural Insights Team's Update Report: 2015-16; http://www.behaviouralinsights.co.uk
[3] Behavioural Insights Applied To Policy, European Report 2016; https://ec.europa.eu/jrc/en/research/crosscutting-activities/behavioural-insights /

marketing or coercive manipulation. Therefore, nudges should be an option that is less intrusive to freedom of choice compared to the other tools in the policy maker's toolbox, apart from education. But education in itself is not an adequate solution (to stay away from undesired behaviors) according to Thaler and Sunstein (2008). Furthermore, it has been argued that since nudges are helping individuals achieve what they want to achieve, nudging is not paternalistic, but rather a form of Libertarian paternalism (Sunstein and Thaler, 2003). Sunstein (2015; 2017a; 2017b) argues that some nudges might be less paternalistic than forcing people to make active choices. These two aspects position nudges as an attractive alternative for policy makers around the world, but what is choice architecture and why would people need help to make good choices?

People need choice architecture for many reasons

If people were perfect decision makers with unlimited time and resources then choice architecture would be ineffective, even useless (Sunstein, 2014). But since people have limited time, resources and cognitive capacity they often use simplifications to base their decisions on (Gigerenzer and Todd, 1999). One way to simplify the decision process is either to use simpler decision strategies (e.g. choose the same as last time or ask someone you trust) instead of more complex strategies (e.g. calculating the utility and probability of each alternative and pick the option with the highest total value across all attributes). Another way to simplify is to decrease the amount of information you base your decision on or to make rough estimates instead complete calculations (Kahneman, 2011). Choice architecture is about helping people to use simplifications in an effective way or to steer away people from using unfit simplifications (according to the choice architect).

It is still a common misconception that people only need more data to make better decisions. It has been recognized in decision science for some time that people do not make optimal decisions and in the 1950s Herbert Simon put forward the concept of bounded rationality (Simon, 1955) to account for the non-optimal behavior. According to bounded rationality, instead of bothering to make optimal decisions all the time, most people take the first alternative that is satisficing. In order to save time and cognitive resources, humans use heuristics and our environment to interpret and navigate our everyday life. Bounded rationality can be thought of as a scissor, where one blade is cognitive heuristics (or limitations) and the other blade is the structure of the environment, and these blades cut the problem space so that a satisficing answer can be found (Simon, 1955). For example, when people are deciding what to eat at a restaurant they usually do not calculate a "price-to-taste-and-calories" ratio (along with all other possible attributes that can be calculated). Instead they apply some kind of decision strategy such as take something I have tried earlier and like, the healthiest looking choice, or simply the first thing they see on the menu that is not too expensive (and looks good). The threshold of what to choose is not determined by which option is

optimal but rather which option is good enough given the time and effort they need (or want) to invest to generate a decision. The closer people are to indifference the less time will be spent to evaluate the different alternatives. Unlike a fully rational agent, humans almost never evaluate all the alternatives and compare all possible outcomes, instead they use heuristics (Kahneman, 2011). Cognitive heuristics (or rules-of-thumb) are effective cognitive strategies that help humans perceive, update and understand information and help people to act on information (Gilovich et al., 2002). The view of humans as fully rational beings (i.e. a homo economicus) has shifted toward viewing humans as homo heuristicus (Brighton and Gigerenzer, 2008).

Homo heuristicus and the two systems

The idea of homo heuristicus, who does not have unlimited cognitive resources and therefore is afflicted by bounded rationality and engages in heuristic use to save the precious cognitive resources at their disposal, is more descriptive of individuals' decision making. The use of heuristics can lead to fast and frugal decisions (Gigerenzer and Todd, 1999), but can also lead to errors and cognitive biases (Kahneman, 2011). Cognitive biases are most often considered to be driven by what Kahneman (2011) calls System 1 (intuitive processing) while debasing is more likely to be considered a System 2 (analytical processing) effect. However, System 2 processing can also lead to biases and System 1 can be used to debias System 2. Kahneman's division of System 1 and System 2 is just one of many dual-process theories. The two-system perspective is not a neurological division of brain regions but rather a perspective of different modes of information processing. The details for the different theories differ but the overall division, as one system/processes as intuitive and the other system/processes as more analytical, is overarching the theories, see Evans (2003) and Evans and Stanovich (2013) for a discussion.

There is no absolute threshold between the two systems, but still they are characterized by certain attributes. The negative attributes of System 2 are that the processes are more resource demanding, slower and System 2 can only process information serially. The positive attributes of System 2 are that it is under conscious control, can handle abstract information, falsify, and use deductive reasoning, and reason about the past and future events. The positive attributes of System 1

are that the processes are very resource efficient, fast and can process information in parallel. System 1 is automatic and is tightly coupled to perception and emotions. But the strengths of System 1 can also be a weakness. System 1 cannot be turned off, it is operating outside of conscious awareness and is stimulus-response driven, which might lead to various cognitive biases. One problem with how System 1 are influencing human information processing. Is that it becomes impossible for people to know which context that System 1 have used to influence the information processes (and how or why). Everyday behavior and decisions are generated from a combination of System 1 and System 2 processes. Since the cognitive system is trying to be resource efficient it tries to unburden the resource demanding System 2 by engaging in cheap System 1 processing as much as possible (Wickens et al, 2015; Kahneman, 2011). This is why we form habits and create heuristics to guide our everyday behavior. However, making decisions or judgements with System 1 is not necessarily better or worse compared to using System 2. This depends on the decision space (e.g. selecting the next move in chess or how to spend your vacation), familiarity, complexity of the problem, emotional information (Slovic and Västfjäll, 2010; Västfjäll and Slovic, 2013; Dickert, Västfjäll and Slovic, 2015; Dickert et al., 2015; Västfjäll et al, 2016) and sometimes is intuitive responses both more accurate and faster compare to more deliberative processing (Klein, 1999). Furthermore, have it been found that fast and slow decision-making affects risk taking (Tinghög et al., 2016; Kirchler et al, 2017; Persson et al, 2018). One important factor for how the two systems might handle decision making is the feedback in the situation. If the feedback is delayed it might be easier to interpret it in System 2 (e.g. to evaluate your long-term saving plan), while instant feedback might be easier to process with System 1 (should more salt be added to the meal?). If System 1 is correctly trained to handle the feedback it is more efficient at reaching a satisficing decision compared to System 2. Both systems are capable of using various heuristics efficiently but also to be error prone at times (Gigerenzer, 2014; 2015), and neither system should be considered the gold standard for making decisions across contexts.

Planner-doer model and consequences

Thaler and Shefrin (1981) tackled the problem with people that misbehave according to economic theory by proposing a planner-doer model for people's decision making over time. The model describes individuals' decision making from an organizational perspective, which consist of two parts, the planner and the doer. The planner part of the model is concerned with lifetime utility (e.g. have money for retirement) while the doer is concerned about here and now (e.g. buying a cool but expensive gadget). The problem is that the planner needs to either impose restrictions or alter the incentives of the doer to make sure that all the resources are not depleted by the doer for short term utility. Similar models have been proposed in economics, e.g. a dual-self model of impulse control by Fudenberg and Levine (2006) with a proposed short-term self versus a long-term self. O'Donoghue and Rabin (1999) argued that a present-bias could explain why people are time inconsistent with their decisions. They suggest that people are either novices, and cannot foresee that they will have self-control problems, or sophisticated, and can foresee that self-control problems will arise in a future situation. Sophisticated people are better at counteracting the present-bias by increasing the (perceived) future cost of a bad behavior now, while novices more often suffer from present-bias. Present-bias occurs when we face either immediate rewards with a delayed cost (e.g. get a high interest loan in order to take a vacation abroad), or an immediate cost with a delayed reward (e.g. exercise to improve your health). These models and the notion of present-bias show that people (some more than others) sometimes need help to achieve long-term goals and avoid bad short-term decisions. It has also been shown that self-control predicts financial well-being in addition to financial behavior (Strömbäck et al., 2017).

This strive of the cognitive system to be efficient or even lazy (Haselager et al, 2008) leads to a constant struggle between the preferences set by System 1 (immediate rewards, present-bias or the doer goals) and the preferences set by System 2 (usually long-term rewards or the planner goals). One occurrence of this could be eating candy instead of exercising and the more long-term preferences of being healthy. Eating the candy instead of exercising might afterwards lead to angst and regret, when you ponder on why you did not follow your (the planner's) preference for staying healthy and promise

yourself that next time will you not fall into temptation but rather follow your preference! However, since System 1 is all about the present and seeks instant gratification you will likely once again switch your (the doer's) preference toward eating candy rather than doing exercise if the choice situation is set up in such a way that eating candy is the path of least resistance, and the exercise can be postponed to another day without (perceived) dire consequences. To further complicate individuals' decision making it has been shown that humans weight their decisions depending on the current reference point.

Reference points and prospect theory

Both the planner and the doer need to construct their goals and evaluate their alternatives somehow. It has been known for quite some time in decision research that reference points and perceived alternatives matter for evaluations for both outcomes and alternatives (e.g. Fredrick et al., 2009; Hsee and Zhang, 2010; Kahneman and Tversky, 2000; Kahneman, 2011). One of the most prominent theories in behavioral economics, prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), tries to explain how people actually make decisions and evaluations (a descriptive theory) rather than how people should make decisions (a normative theory). The two main contributions of prospect theory are that the perceived (or felt) gain or loss depends on your current reference point and that a proportionally equal gain (e.g. winning 100 dollars in a lottery) has a smaller relative value compared to a proportionally equal loss (e.g. losing 100 dollars in a lottery). The relative value of money (or any other utility) always depends on your current reference point and the just noticeable difference (JND) factor. In other words, winning 100 dollars makes you happy, but you do not get happier for winning 101 dollars compared to winning 100 dollars (as long as your JND is bigger than 1%). The idea of importance of the reference points in prospect theory comes from perceptual adaption (Kahneman 2003), and the concept of just noticeable difference comes from the Weber's law in psychophysics and is being used to measure human perception (Fechner, 1860; Weber, 1834). Kahneman and Tversky (1979; Tversky and Kahneman, 1992) argue that this relativism is prominent in all judgements and evaluation humans do and that nothing is valued in an

information vacuum. Unstable preferences that can be constructed on the spot are why nudging or choice architecture sometimes is needed and the way we process information is why it can work at all. It is often said that change comes from "the inside", but sometimes (more often than most people believe), the means to change have to come from "the outside" which is at the core of nudging.

The six principles of nudging

The intent of nudges was to use six principles of good choice architecture to help individuals make better decisions by avoiding mental missteps and biases. Nudging could in theory help individuals avoid any kind of biases. In their book *Nudge: Improving Decisions about Health, Wealth, and Happiness* (2008), Thaler and Sunstein give a few example of biases that are important to nudging: Anchoring (occurs when you use a reference point to adjust your estimate of a value; Tversky and Kahneman, 1974), Availability (the ease with which an occurrence can be brought to mind is used to assess the probability of that event; Tversky and Kahneman, 1974), Representativeness (probabilities are evaluated by the degree to which A is representative of B; Tversky and Kahneman, 1974), Optimism bias (that people believe they are less likely to experience a negative outcome compared to others; Sharot, 2011) and Overconfidence (overestimation of one's actual performance; Moore and Healy, 2008), Gains and Losses (a loss of an amount seems larger than a gain of the same amount; Tversky & Kahneman, 1991), Status Quo (people tend to weight the potential losses of switching from the status quo more heavily than the potential gains of switching; Kahneman, Knetsch and Thaler, 1991), Mental accounting (people divide their assets into different mental accounts; Thaler, 1999) and Framing (people react differently to the same information depending on how it is presented; Tversky and Kahneman, 1981). These biases can induce semi-optimal behavior and can be debiased with nudging (or induced with nudging for that matter). The six principles for good choice architecture are, according to Thaler and Sunstein (2008), incentives (which should be salient), understanding mappings (help individuals to map and therefore choose options which are good for them), defaults (make sure that the pre-selected option is a good one), give feedback (people make mistakes, and a good design takes that into account), expect error (to

err is human and a good design helps people recover from their errors), structure complex choices (provide strategies and structure information to help individuals interpret the situation) or:

iNcentives
Understanding mappings
Defaults
Give feedback
Expect error
Structure complex choices

There are two potential ways that nudges, or choice architecture, could be used to change behaviors, either by debasing the decision maker or to structure the situations so that the decision maker will be likely to use a specific heuristic which is deemed as beneficial in that situation. However, neither biasing nor debiasing individuals is problem free.

Nudge techniques

The idea of using choice architecture, or nudging, to help individuals to make better choices (as judged by themselves) has not escaped criticism. Nudging has raised several ethical concerns where the first question to answer is whether nudges are paternalistic and if so, what are the boundaries for using nudges (e.g. Schubert, 2017; Sunstein, 2014; Sunstein, 2016b; 2017a). Nudging as a concept has evolved and different ways of categorizing nudges along the cognitive spectrum have emerged and various nudge techniques are suggested in the literature (e.g. default-changing, reminders, emotion-eliciting). The simple model presented in Figure 1 consisted of three parts were part 1 is nudge techniques. This part can be divided into several components since nudge techniques can vary over several factors.

    As can be seen in Figure 2 has nudge technique been expanded to include *Type* of nudge (since Type 1 and Type 2 nudges use different types of psychological mechanisms to affect behavioral change), *Transparency* (since non-transparent nudges threatens people autonomy), *bias or debias* (since this strengthen or weaken people rationality), *social* nudge (and engage more social motives instead of more cognitive or emotional information processing) and *ease of use* (if it will cause friction for individuals when it is implemented or not).
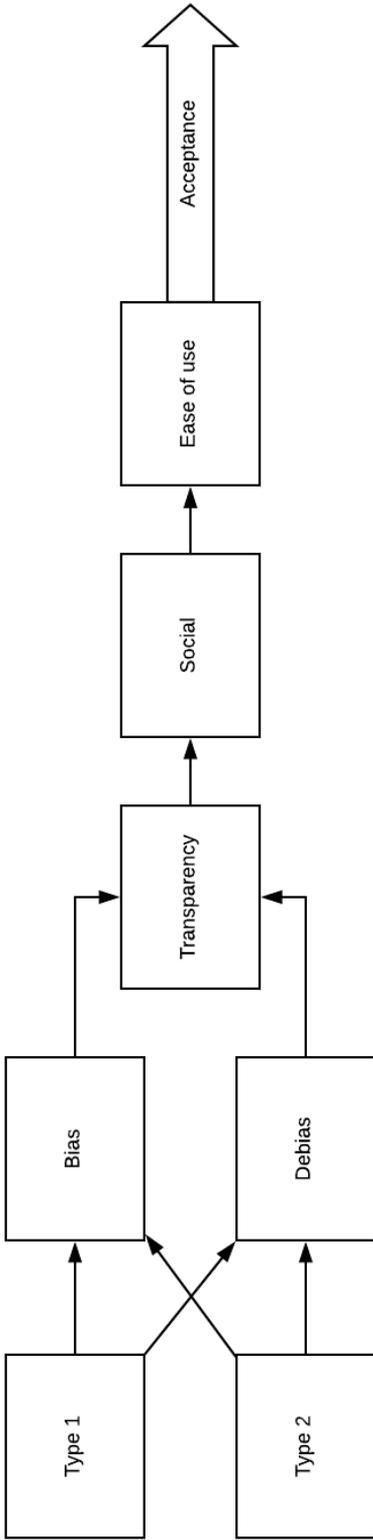
Figure 2. Model of Nudge technique, showing different components for affecting psychological mechanisms that leads to behavior change.

Type of nudge

Hansen and Jespersen (2013) divide nudges with regards to two dimensions, the transparency dimension (e.g. level of transparency) and the Type of nudge dimension (i.e. Type 1 or Type 2). Type 1 corresponds to nudges that engage System 1 processes and Type 2 engages System 2 processing. In other words, a Type 1 nudge is a nudge that is hard to distinguish from perceptual cues, associations and automated processes while a Type 2 nudge utilizes the reflective System 2.

Hansen and Jespersen argue that Type 1 nudges are targeting behavior to a greater extent instead of choice (although this depends on how a choice is defined), since a well-functioning Type 1 nudge could affect an individual without the individual realizing that a choice even has been made (especially if the nudge also is non-transparent). Nudges that are operating on the level outside of consciousness and shape choices for the individual without the opportunity to object are easily considered paternalistic, since there can be no freedom of choice without a choice at all.

Hausman and Welch (2010) suggest that nudges (that could be categorized as Type 1 non-transparent nudges) are very intrusive and might be more paternalistic compared to bans or regulations, since a choice architect exploits cognitive imperfections of an individual to affect the individual's judgment in a given context. However, they find nudges less problematic in those cases when individuals themselves can set up nudges (even Type 1) in order to increase their ability to choose rationally at a later time (e.g. add themselves to a ban list from casinos). Many Type 2 nudges could be perceived as rational persuasion and should therefore not be considered to be paternalistic at all according to Hausman's and Welch's reasoning.

Baldwin (2014) takes a similar stance and argues that the more impact nudges have on individuals' autonomy the more problematic they become. Baldwin divides nudges into three degrees. The first degree is unproblematic and is simply prompting correct information to individuals at the moment of choice (which could be considered a Type 2 nudge). The second degree is more problematic and is more infringing on the autonomy of individuals with default options or design of a public area to promote (or negate) a certain behavior. The

third level is also infringing on autonomy (and freedom of choice) but in a more devious way, which is hard for the individuals to safeguard themselves from or even notice. Examples of the third degree are emotional stimuli or other non-transparent Type 1 nudges. Baldwin states that it is hard to define exact conditions for nudging to make the nudges efficient and yet acceptable. Furthermore, Baldwin argue that as efficiency of nudges increase might the acceptability decrease, especially for nudges which are not transparent.

Transparency

The dimension of transparency, according to Hansen and Jespersen (2013), is more focused on whether the person who is being nudged can reconstruct both the intention behind the nudge and the means of the behavioral change. With a truly non-transparent nudge, the person being nudged will not notice the nudge at all (and sometimes even that a choice has been made). From the cognitive perspective of nudging (how it works) transparency does not necessarily matter so much, but from the ethical perspective it might be a key factor that separates acceptable nudges from coercion. Bovens (2009) argues that for a nudge to be sufficiently transparent, it needs to pass the "*In principle token interference transparency*" requirement. To pass this requirement, the intention of each token (i.e. the individual application of a nudge in a given context), in principle, should be identifiable (and comprehensible) by a watchful individual with reasonable effort. In other words, it is not necessary that every nudge is identified and understood by all individuals that are being nudged or every time an individual are being nudged. Schmidt (2017) argues that this kind of transparency is a necessary, but not sufficient, condition to accept nudges. In order for nudges to be acceptable they also require democratic control (e.g. citizens need to have a way to accept or reject the usage of nudges).

The arguments presented so far are favorable for transparent Type 2 nudges as rational persuasion and maybe Type 1 nudges if they are transparent. But that raises the problem of how transparency can be guaranteed for nudges. Lepenies and Malecka (2015) propose that this could be achieved by complementing nudges with legal norms so that they connect with the legal system. Making nudges part of the legal system has two advantages according to Lepenies and Malecka. First,

it makes them public and it gives people a chance to oppose the usage of nudges in different domains. Second, it makes it possible to create a nudge registry over acceptable nudges and a "nudging ombudsman" that could be appointed by parliaments to meet legal concerns and to ensure that nudges conform to constitutional and basic legal principles. By making nudges part of the legal system and therefore transparent and possible to reject, part of the problem with nudges could be fixed according to Lepenies and Malecka.

Bias or Debias

Regardless of type or transparency, nudges can be used either to debias people from a behavior or to bias people to engage in a behavior. The (de)biasing can be both emotional and cognitive. A cognitive bias is most simply described as a mental misstep of some kind. An evolutionary taxonomy of cognitive biases proposes that there are three type of biases, heuristic biases, error management biases and artifacts biases (Haselton, Nettle and Andrews, 2005). Heuristic biases are either when we use an efficient heuristic in the wrong domain or when a situation is misinterpreted, and we therefore "select" an erroneous heuristic (e.g. pour water on an electrical fire). Error management biases are when the decision error is weighted so we are more likely to err in the less costly direction (e.g. people tend to fear snakes even in countries where no snakes are dangerous). This can be used with nudges to either rig a situation to avoid the biased behavior or set up the choice situation to maximize the likelihood of people using a certain kind of heuristic, which will generate the wanted behavior. Another way to affect people's behavior is to play on their emotions. Our emotions or the affective properties of an object can be used to guide our attention and weight information (e.g. Clore and Huntsinger, 2007; Huntsinger and Schnall, 2013; Huntsinger et al., 2014). With nudges you could either try to debias people by changing their emotional states or bias people into an emotional state to affect their decision making. The use of nudges to debias people should be less problematic from a paternalistic or ethical point of view since it might increase (rather than decrease) the autonomy of the individual. However, to use nudges to bias people into a certain behavior (even if it is preferred by the individual) is more questionable from an ethical perspective since it decreases the individual autonomy. Another similar

suggestion is increasing people's choice capability with boosts rather than nudges (Grüne-Yanoff, and Hertwig, 2016). A boost is argued to increase autonomy by increasing individual's decision making by boosting their decision technique rather than exploiting psychological mechanisms (Hertwig and Grüne-Yanoff, 2017).

Gigerenzer (2015) argues that instead of nudging people, the focus should be to educate people in statistical and critical thinking and to be understanding of risk in general. Furthermore, how can we be sure that the choice architects themselves do not suffer from various cognitive biases when they decide on what, when and who to nudge? They could easily fall into the same trap that they are trying to prevent everyone else to fall in. Lodge and Wegrich (2016) note that it is indeed ironic that nudging, which is dealing with reasoning failures and bounded rationality, has so little self-awareness of its own limitations. Unsurprisingly it has been found that public employees also are subjected to cognitive biases (Bellé et al., 2018). Furthermore, elected politicians are equally, or even more likely, to fall for certain cognitive biases as the general public (Sheffer et al., 2018). Another approach to change people's behavior could be to target social motivations with nudges.

Social motives

According to Nagatsu (2015), social nudges are not problematic from an autonomy point of view, since they are more reason based and easy to avoid. A social nudge is, for example, presenting what the typical behavior is in a given context (e.g. a message sent out with your tax declaration saying, "Most people pay their taxes on time in your community"). This reference point can then be used to guide individual behavior (Furth-Matzkin and Sunstein, 2017), it might also change the framing of the problem from an I-frame (individual point of view) into a We-frame (as part of a group or community) which might impact the decision at hand but it does not, however, disable reasoning and should therefore be an acceptable nudge even with a pro-social goal. Using social norms as a nudge might be effective for some groups but not for others. Schultz et al. (2007) showed that stating the social norms can produce a boomerang effect (people act in the opposing direction instead) for the people that are already outperforming the stated norm. They argued that the impact will depend on the distribution of the

behavior in the population. Mols et al. (2015) proposes that one way to increase the effectiveness of social influence[4] and also turn it into a lasting behavior change is to make nudges that help individuals to internalize social norms rather than just adapt to the norm. For this to happen is it important that people feel that they share a social identity with the messenger and therefore make the norm an integral part of their self-understanding.

As a response to highlight this (implicit) interaction between the individual being nudged and the choice architect, Krijnen et al. (2017) have proposed a new conceptual framework, which they call choice architecture 2.0. They argue that people engage in social sensemaking when faced with a decision which is divided into two parts. The first part is the individual's interpretation of the beliefs and intentions of the choice architect. If the individual perceives the choice architecture intentions to be coercive or a threat to their freedom of choice then they will try to counter the intended effect (e.g. see theory of reactance, Brehm and Brehm, 2013). The second part of choice architecture 2.0 is the meta-cognitive reflection on how others (including the choice architect) will interpret their response to the nudge. For instance, people might follow a default rule to signal that they understand that this is the appropriate thing to do or choose a different alternative to signal that they have thought about the subject. In other words, people can use their behavior as a signaling strategy to communicate to others what they consider to be important. This can either be exploited with nudges or become problematic when implementing nudges.

Ease of use

People tend to like change in general, but do not like to change their own behavior, due to several biases (e.g. status quo, Omission bias, endowment effect). The best nudge is a nudge that people do not perceive to be cumbersome or annoying, unless the point of the nudge is to reduce a behavior. The ease of use of nudges is not really a problem from the ethical perspective but rather a functional factor that may make the nudge an annoying disturbance rather than a functional

---

[4] Mols et al (2015) *differentiate social influences from nudges, since they target social identity rather than cognitive processes of individuals which Furth-Matzkin and Sunstein, (2017) and Nagatsu (2015) among others do not.*

decision modifier. There are various reasons why people might say they want change (saying so costs nothing) but the action to make (or to engage in) the change is missing. For instance, people are resistant to change because of status-quo and omission biases (Tversky and Kahneman, 1974; Ritov and baron, 1992), but also because of the endowment effect and loss aversion (Kahneman, Knetsch and Thaler, 1991). These mechanisms make us overvalue the situation we are in compared to change and also make us feel worse off if the outcome of the change is worse compared to the previous state. This intrinsic cost of change makes it a first threshold to behavioral change. The other threshold is how easy it is to engage in the behavior. In technology usage it is known that ease of use (or effort expectancy) is an important factor (along with perceived usefulness) for people to take up and use new technology (e.g. Davis, 1989; Davis et al., 1989; Venkatesh et al., 2003). Even though it has been challenged how big the effect of ease of use is on technology adaption, I argue that it should be an important factor for effectiveness of nudges (i.e. people will resist increased hassle if they can). One problem with comparing how much impact ease of use has on adapting to new technology with the efficacy of nudges in a workplace is that people tend to not have much choice on whether to use new technology or not, regardless of initial learning costs. A nudge on the other hand should (according to its definition) be easy to avoid, and therefore is it important to make nudge to be cheap in initial learning costs.

Using nudges to change behavior

Even if nudges can be questioned from an ethical point of view they are probably here to stay in one form or another. Mainly because they seem to work in various forms and contexts (e.g. Milkman et al, 2011; Kalbekken and Sælen, 2013; see, Szaszi et al., 2018 for a review) and are cheap and easy to implement. The effect sizes vary but for many nudges (e.g. increase repayments, Cadena and Schoar, 2011) small effects can have a great impact when the number of affected people increases. Defaults seems to have a great impact on choices and Johnson and Goldstein (2003) argue that a simple switch of the defaults from opt-in (an active choice needed to become a donor) to opt-out (an active choice is needed to stop being a donor) could substantially increase the number of organ donors.

But nudges are no silver bullet, they do not always work (e.g. de Wijk et al, 2016; Sunstein, 2016b; 2017c; Kosters and Van der Heijden, 2015), sometimes effects have been overstated and the effects vary over demographic groups (e.g. socio-economic status) and the behavior is increased for people that already are participating in the behavior, while others are unaffected (Rivers et al, 2017). The effect of nudges seems also to vary over political orientation, Costa and Kahn (2013) found that liberals reduced the electrical consumption to a greater extent compared to conservatives and were also less likely to drop out of the experiment.

Goodwin (2012) argues that nudges, in addition to being troublesome in relation to freedom of choice, are not effective enough to solve society's major problems. Furthermore, he suggests that the problem with nudges is that if you can nudge people in one direction (e.g. eating healthier) then they should also be vulnerable to be nudged back again by marketing campaigns.

Kosters and Van der Heijden (2015) suggest that another problem with nudges is that it is a bit unclear how to evaluate nudges and that the definition of what constitutes a nudge is also rather fuzzy, which adds to the problem. For instance, is the nudge to be considered a success if it increases tax compliance with 1% or increases the average amount of carrots people eat by 50%? Koster and Van der Heijden (2015) also point out that some of the success stories of nudging might not be as great as they look on first glance, but they still have positive

effects. The lack of a solid evaluative framework makes it easy for people to be either over enthusiastic or over pessimistic about the effects of nudges and how and when the results are generalizable (since nudges are often assumed to be context dependent). Furthermore, Koster and Van der Heijden (2015) conclude that nudges are often an addition to, not a replacement of, other policy tools. This, while positive for all the nudge sceptics, makes the net effect of the nudge harder to evaluate. Barton and Grüne-Yanoff (2015) propose that in order to get support (and normative justification) for nudges it is crucial that evidence for the psychological mechanism of real-world nudges is confirmed and understood. Several attempts to create guidelines for nudges have emerged (e.g. Ly and Soman, 2013; Dolan et al, 2012) in order to increase the efficiency of nudges, but so far little attention has been paid to which factors affect the acceptability of nudges.

Nudges, paternalism, ethical problems and why acceptance matters

Since nudging was introduced the concept has been debated and questioned on several important aspects. For instance, why should nudges work at all and which different kinds of choice architectures could be created (e.g. Thaler and Sunstein, 2008; Hansen and Jespersen, 2013; Sunstein, 2014; Grüne-Yanoff, 2016)? What kind of effects could nudges be expected to generate and when should they be used (e.g. Felsen and Reiner, 2015; Sunstein, 2016c; Sunstein, 2017a)? Are nudges paternalistic in their nature and if so is that equally problematic for all kinds of nudges (e.g. Baldwin, 2014; Hansen, 2016)?

This raises two problems for nudges. The first problem is that people can have different preferences and therefore it is problematic to have general nudges. This could, however, be tackled by using individual nudges according to Mills (2015). If (and only if) the nudges are individually designed (in accordance with the individual's own preference) then the nudge would not necessarily undermine personal autonomy. Personalized nudges might even strengthen personal autonomy by helping individuals to not sway from their chosen path (or focus their energy on other things), but even then, is it important that the individual can chose to opt-out of the nudge and

make other choices. This in turn generates a privacy concern according to Kapsner and Sandfuchs (2015), since each individual's preference needs to be extracted in some way (including the preference of not having your preferences extracted) to be able to create personalized nudges. Another problem is then how to decide on how this data should be permissible to use for each individual. This also raises the question of how often, when and how should the individual preferences be updated?

This leads us to the second problem, which is whether people even can have stable preferences. Since the struggle between the preferences of System 1 and System 2 is one of the reasons nudging should work (or even be needed at all), it seems strange to argue that individuals can hold stable preferences at all. If they can, how can one preference be valued over others?

Unstable preferences

Humans can hold individual preferences and these preferences govern the decisions we make. Economists argued for a long time that preferences are stable (and well-defined) and can be revealed by observing individuals' actions or asking them for their preference. By asking people if they choose goods X over goods Y, the choice of goods X implies that goods Y are less preferred and therefore are considered inferior to goods X (Samuelson, 1938).

The idea of stable preferences has since then been challenged and the view that preferences can be constructed in the process of elicitation and are dependent on contextual factors (Lichtenstein and Slovic, 2006) has emerged. The idea of preference construction explains why people can prefer different things at different times, and why people behave in a way that goes against their previously stated preferences. It has also been shown that preferences can be reversed depending on evaluation mode (Bazerman et al., 1992; Hsee et al, 1999).

Whitman and Rizzo (2015) sum up what they call the problem for the behavioral paternalist (advocates for nudges) with:

> The paternalists use evidence of internally inconsistent preferences as evidence of irrationality – and then, with little or no justification, they select among the inconsistent preferences to determine what someone's "true" preferences must be. (p 410)

The problem as they point out is that there is no normative standard to judge which preferences are to be considered the "true" (or stable) preferences of the individual. If you try to extract the preferences by asking individuals, they might (most probably) give you a biased answer. Withman and Rizzo (2015) argue that the behavioral paternalists have accepted the neoclassical axioms for rationality (well-defined preferences) as the standard for individual welfare. In other words, within everyone is a latent homo economicus that is being suppressed by a homo heuristicus. This sounds like a rehash of the debate if System 2 is a better way to make judgment and decisions compared to System 1, which is both context, domain and experience dependent (Kahneman, and Klein, 2009). This is further complicated by the claim that the systematic biases are neither that systematic nor need to be irrational behavior per se (Gigerenzer, 2015). One alternative approach to explain some of the logical errors people make in experiments is that the errors are actually produced by the use of strategic interaction or social intelligence (which usually works fine outside of the lab but might work poorly in a lab setting).

Why some nudges fail

Nudges could fail because of three major reasons. The first reason is that the problem the nudge is supposed to help avoid is misunderstood and what the choice architect is trying to accomplish with the nudge is simply not the solution to the problem. The second reason is that the design of the nudge is flawed and does not enable (or disable) the psychological mechanism needed to change the behavior (i.e. poor nudge technique). For instance, adding a small fine for parents that were late to pick up their kids from daycare centers increased the frequency of parents showing up late instead of decreasing it (Gneezy and Rustichini, 2000). The third reason is that the person being nudged finds the nudge to be an unacceptable manipulation.

Lehner et al. (2016) point out that some nudges seem to work better than others and acceptance of the nudge might be a contributing factor. Therefore, it might be valuable to measure the acceptance for nudges in the population that is going to be the target of the nudge. Furthermore, nudges have backfired when the goal or the choice architect has been questioned by the people being nudged. Switching the default for organ donations (to a presumed consent legislation)

actually decreased the donation rates in Chile (Dominguez and Rojas, 2013) and several attempts to switch the defaults in the US have had to be abandoned since they lack acceptance from various large groups (Bard, 2012). These effects could be explained by the theory of reactance. Reactance is a motivational state that occurs when the individual perceives their freedom is threated. The magnitude of the reaction depends on the level of the threat and also on how important the specific freedom is to the individual (Brehm and Brehm, 2013). Reactance is a common phenomenon in various context and not just with nudging (Rains, 2013).

Using a social nudge (e.g. stating the norm) on people with a firm belief in a subject can result in reactance and non-conformity (Furth-Matzkin and Sunstein, 2017), if the nudge goes against their view. This is most likely true for more than social nudges, when people being nudged feel that the goal (or the means) of the nudge goes against their firm beliefs. The magnitude of reactance usually increases if people feel that information is presented in an attempt to persuade them or others, (Brehm and Brehm, 2013). In contrast, people without a firm belief are more likely to be nudged. However, the nudge literature has had little focus on what makes nudges acceptable or not by the general public. To complement this theoretical discussion about whether nudges should be acceptable or not, I here present empirical data from the general public.

Aims of the thesis

The overarching aim of this thesis is to present a model with concepts that influence the acceptance of nudges, and to empirically measure the influence of those factors by systematically varying attributes of the nudge (nudge technique, goal of the nudge and stated beneficiary) and factors around the nudge (alternative to the nudge, and who is doing the nudging). This thesis will furthermore explore if nudges categorized as pro-self (more beneficial for the person being nudged compared to society) are more acceptable compared to pro-social nudges (more beneficial for society at large compared to the individual being nudged).

Moving from a theoretical debate into measuring acceptance – including summary of empirical papers

If the general public would find nudges to be unacceptable, then it would not matter whether nudges are theoretically intrusive to freedom of choice or not. On the other hand, if people could accept nudges then it would be interesting to examine the boundaries of that acceptance. Furthermore, it is interesting to investigate whether a nudge could be considered intrusive to freedom of choice and still be acceptable or if that is two mutually exclusive factors. As of yet, the question of whether nudges are acceptable or not has mostly been a debate among academics on a theoretical level, while the policymakers are more likely to be interested in empirical results which can specify when a nudge is acceptable or not by the general public in order to implement them. The acceptance of the people that are being nudged is important to fulfill the condition of democratic control.

In this thesis will I try to increase the understanding of when and what makes a nudge acceptable by presenting a model with factors that could affect acceptance and then empirically testing these factors in a systematic way. Many frameworks have been presented for functionality of nudges and how they could be designed, e.g. MINDSPACE (Dolan et al., 2012), Choice Architecture 2.0 (Krijnen et al., 2017) But these frameworks are focusing on the first part of the model for behavior change with nudges, the nudge technique (see Figure 1), but not on the second equally important part, acceptance of nudges.

Method

The empirical data on acceptance in all four papers were collected with online web questionnaires in order to be able to get representative samples. The questionnaires started out with introductory information about what the questionnaire was about, followed by the scenarios. The scenarios described in all surveys had the same basic structure (with varying details for each paper depending on manipulations and contexts). That structure consisted of first presenting a problem (e.g. there is a lack of organ donors in Sweden today), secondly presenting a

suggested policy intervention (e.g. to make it easier for people to become organ donors by changing the organ donor registration procedure to opt-out instead of opt-in) followed by a description of the intended consequences of that intervention (e.g. this switch will increase the number of organ donors in Sweden). In paper 1 acceptance, intrusiveness to freedom of choice, and benefits for individuals or society at large were measured with a 1 to 4 scale. This was change in paper 2 two so that acceptance and intrusiveness were instead measured with a binary yes or no answer, while benefits were still measured on a 1 to 4 scale. Paper 3 and 4 used the binary choice for acceptance and intrusiveness while the benefit questions were changed to a -3 to 3 scale in order to capture whether individuals found any negative impact of the interventions. There are several ways to measure attitudes towards nudges (or interventions in general) and we wanted to keep the measure as similar as possible between the studies. We settled on acceptance since it captures the lowest levels of tolerance needed in order to let something pass. People can dislike something and be unsupportive of it yet at the same time accept it, much like most people do not like to pay taxes, but they accept that they need to in order for society to keep functioning the way it is. Acceptance is a concept which includes more than mere liking and at the same time captures when an intervention passes the boundaries of what individuals perceive to be tolerable to do. At the same time individuals could signal with the intrusiveness question and benefit ratings whether the intervention (nudge) was considered good or bad along those dimensions.

Pro-social nudges and pro-self nudges

One of the core attributes of nudging according to Thaler and Suntein (2008) is that the nudge should guide individuals to engage in behavior which is beneficial to them, judged by the individuals' own preferences and not by the preference of the choice architect. However, far from all the nudges have only the benefit for the individual in mind. I make a distinction between pro-self nudges (beneficial to the individual being nudged) and pro-social nudges (beneficial to society at large). Pro-self nudges are more often designed to help individuals act more in line with a rational agent (increasing the individual's utility) and help individuals tackle cognitive biases that generate

irrational behavior. Therefore, making the individuals act more like a homo economicus and less like a homo heuristicus. Pro-social nudges, on the other hand, can promote irrational behavior and induce cognitive biases that generate a behavior which makes individuals act less like a homo economicus and more like a homo heuristicus which might decrease the individual's utility (while increasing society's utility).

The goal of theses nudges is fundamentally different even if most nudges could be viewed as beneficial for both the individual and society at large, but a pro-self nudge could potentially be costly for society while a pro-social nudge could generate costs in some form for the individual being nudged. Pro-social nudges cannot be defended on the same ethical grounds as pro-self nudges, and a Type 1 non-transparent pro-social nudge is always paternalistic while a Type 1 pro-self nudges might not be, if (and only if) the nudge actually helps the individual engage in behavior along the individual's (stable) preferences.

One alternative given by Guala and Mittone (2015) is that we abandon the welfarism idea of nudging (i.e. what I call the genuine pro-self nudges) since it is a red herring. Nudging is cheap and more liberal (often not always) compared to bans and regulations. But there is no point in labeling it as a pro-self nudge, when most of the nudges that are being implemented have a pro-social agenda, like environmentally friendly nudges (e.g. Schwartz et al. 2012) or to increase tax compliance (e.g. Hallsworth et al, 2017). Nudging is a tool that can be used to avoid tragedy of the commons (Hardin, 1968), i.e. depleting the common resource pool for individual benefit, but at the same time not infringe too much on people's freedom of choice.

Paper 1, empirically exploring pro-social nudges and pro-self nudges

In order to investigate if there is a difference in acceptance for pro-self and pro-social nudges, eight common nudges from the nudge literature were selected (see paper 1). For each of the eight nudges the problem, the nudge and the intended goal with the nudge were presented in a vignette. After each vignette the participants rated acceptance, intrusiveness to freedom of choice, benefit for individuals and benefit for society at large. After the participant had answered all eight

vignettes (the order was randomized for each participant) they answered the short version of the culture and cognition worldview scale (CCWV, Kahan, 2014) and a short version of the Rational-Experiential Inventory (REI, Norris et al., 1998) scale. The CCWV scale was used in order to measure whether more individualistic (compared to communitarians) were less accepting of nudges and if people with more egalitarian (compared to more hierarchal) values were more accepting of nudges. The REI scale was used to measure whether individuals' preferred decision styles, either more rational or intuitive, have an impact on acceptance toward nudges. The data was collected online and the participants were recruited from a representative subject pool in Sweden (n = 514) and the US (n = 438). The selected nudges were *a priori* divided into pro-self nudges and pro-social nudges.

Our findings from study one shows that although nudges in general have a high average acceptance, the notion of one-nudge-fits-all is not tenable. There was also a difference in acceptance and intrusiveness between the two countries. Moreover, people with a more individualistic worldview are less likely to accept nudges and people prone to analytical decision style are less likely to judge nudges as intrusive to freedom of choice. Furthermore, there was a difference between the nudges categorized as pro-self compared and the nudges categorized as pro-social. We found that the pro-self nudges have a higher acceptance rate compared to the pro-social nudges and are also less likely to be judged to be intrusive to freedom of choice.

Adding pro-social and pro-self to the model

Although there is much to further investigate from the first paper, the results show that nudges could have a place in the policymaker's toolbox and that the division of pro-self and pro-social nudges have some merit when it comes to acceptance of nudges. Therefore, the attributes of pro-social and pro-self is added to the acceptance part of the model in Figure 3. It is, however, unclear if the difference in acceptance of nudges is partly because of the nudge technique, the goal that is trying to be achieved or a combination of both.

Pro-self

Pro-social

Figure 3. The nudge acceptance model expanded with the first two components that influences acceptance of nudges, benefits for the individual (pro-self) and the benefits for society (pro-social).

Separating goal of the nudge from nudge technique

Even though we learned a lot regarding the acceptance of nudges in the first paper, several questions remain unanswered. Several studies that followed the publication of the first paper have measured acceptance in the general population across different countries and found that a majority of the proposed nudges are deemed acceptable (e.g. Jung and Mellers, 2016; Reisch, Sunstein and Gwozdz, 2017; Sunstein, 2017a; Sunstein, Reisch and Rauber, 2017; Loibl et al., 2018). It also seems that the judgments of nudges are affected by individual differences and attitudes (e.g. Paper 1, Sunstein, 2017a). But what makes the nudge acceptable or not by the general public? Is it the nudge technique (if and how the nudge affects individuals' autonomy and freedom of choice), the goal of the nudge (pro-social vs pro-self) which has been discussed above or is it something else, for instance the domain in which the nudge operates (e.g. environmental, savings, health), or all the above?

Paper 1 showed that nudges as a general concept could be usable for policy makers without enraging the general public. Even though

each nudge was presented in a similar manner, presenting the problem followed by a suggested intervention (nudge) and the intended goal with this intervention (nudge) the nudges varied on several factors making it hard to draw conclusions on which variables are important for acceptance of nudges.

This problem is shared by many of the following studies that investigate attitudes toward nudges (e.g. Jung and Mellers, 2016; Reisch, Sunstein and Gwozdz, 2017; Sunstein, 2016; Sunstein, Reisch and Rauber, 2017; Loibl et al., 2018). For instance, it is hard to compare using a default-changing nudge technique in the context of organ donation (health domain) with an emotion-eliciting nudge technique in the context of decreasing littering (environmental domain) since they both vary on the domain, the intended goal of the nudge and the way the nudge technique is intended to achieve the goal. One current problem in the nudge literature is disentangling the used nudge technique from the context it is being used in and the intended goal of the nudge. For instance, people might accept changing from an opt-in default to an opt-out default in the context of environmental footprints, but not for organ donation. However, if each nudge technique is tested within a unique context it becomes impossible to conclude whether people oppose the nudge technique or the intended goal of the nudge technique.

Sunstein (2017a) proposes that nudges get rejected when the means of the nudge is regarded as manipulative (e.g. a non-transparent Type 1 nudge technique). A few studies have clustered different nudge techniques into categories in accordance with Hansen and Jespersen's (2013) Type 1 and Type 2 nudges and showed that more intuitive compared to reflective nudges are more acceptable and that more transparent nudges seem to be more acceptable compared to nontransparent nudges (e.g. Felsen et al., 2013; Jung and Mellers, 2016; Sunstein, 2016; Sunstein, Reisch and Kasier, 2018). Sunstein (2017a) argues that everything depends on the goal of the nudge and that the nudge technique is irrelevant. However, few studies have measured acceptance of different nudge techniques within the same context and with the same goal. Therefore, it is hard to draw conclusions about whether the nudge techniques affect acceptance or not.

With this in mind we set out to explore three things with paper 2. First, do different nudge techniques with the same intended goal vary

in acceptance compared to each other? Secondly, do people find the same kind of nudge techniques acceptable in different domains? Thirdly, given that pro-self nudges are perceived as more acceptable compared to pro-social nudges is it possible to increase the acceptance of a nudge by framing the nudge as pro-self (explicitly stating that the intended beneficiary is individuals) instead of pro-social (explicitly stating that the intended beneficiary is society at large)?

Paper 2, empirically exploring the effect of nudge technique on acceptance

In order to further explore the parameters of the acceptability of nudges we conduct two new online survey in Sweden (n = 677) and the US (n = 790), see paper 2. In order to keep the design as systematic as possible, the structure from paper 1 was kept for all nudge scenarios. First, the problem was presented with an intended intervention (the nudge technique) followed by the intended goal of the intervention. The nudge scenarios consisted of three different nudge techniques in three different contexts. The nudge techniques were chosen to represent Baldwin's (2014) three degrees of nudges, the degrees vary in intrusiveness to autonomy of the individuals that are being nudged (increases with the degrees). Information-providing nudges were selected to match the first degree. Default-changing nudges were selected to match the second degree. Emotion-electing nudges were selected to match the third degree. A version of each nudge technique was presented in three domains. The first domain was the health domain (decrease smoking), the second domain was wealth (increase retirement savings) and the third was environmental (decrease energy consumption). In order to find out if merely framing the nudges as pro-self instead of pro-social could affect the acceptability of nudges two versions of each nudge scenario were created for the survey. The surveys had a between-subject design where half the participants read nine nudge scenarios framed as pro-self (explicitly stating benefits for individuals) and the other half read the same scenarios framed as pro-social (explicitly stating benefits for society at large).

From the results we learned that all nudge techniques varied in judged acceptance and intrusiveness to freedom of choice for all domains. The information-providing nudges had the highest

acceptance (and most stable over domains), followed by the default-changing nudges with the emotion-eliciting nudges as least accepted. However, the default-changing nudges were more likely to be considered intrusive to freedom of choice compared to the emotion-eliciting nudges, with the information-providing nudges as least likely to be considered intrusive. Furthermore, nudges were more acceptable in the health domain (i.e. decrease smoking) compared to both the environmental domain (decrease energy consumption) and the wealth domain (increase retirement savings) which was least accepted of the three. However, for intrusiveness no overall difference was found between the environment and health domain, but the wealth domain was more likely to be considered intrusive compared to the other two.

No general support that framing (i.e. presenting the information with different focus) could increase acceptance was found, but pro-self framed nudges were more acceptable in the wealth domain for both the information-presenting and the emotion-eliciting nudge techniques compared to the pro-social version of the same nudge. In the Environmental domain the pro-self version of the emotion-eliciting nudge technique was less likely to be accepted compared with the pro-social version. However, pro-self nudges were generally less likely to be rated as intrusive to freedom of choice compared to the pro-social versions.

These findings have some implications for nudges in general. First, it shows that more than the intended goal of the nudge matters. People care about how the intended goal is going to be reached as well as the goal. Secondly, it again shows that even if a nudge is considered intrusive, the nudge can be considered an acceptable approach to achieve certain goals. This might be comparable with legislation where people are willing to forgo some freedom of choice for the greater good and a working society. But people probably need to find the intrusiveness-effectiveness trade-off of the nudge (or legislation for that matter) to be good enough. This goes in line with Petrescu et als., (2016) finding that people's perceived effectiveness is a predictor of the acceptability of nudges. However, only highlighting the intended beneficiary with framing seems to lead to different results in different contexts. The reason for the different effects of the framing could be that the domains themselves have an interaction effect with the framing. Some domains could be considered more or less prosocial in general, for instance environmental questions compared to wealth

creation. If this is the case, then promoting a prosocial domain with individual gains might be counterproductive and yield a negative effect, in a similar way that paying people to give blood decrease some individuals will to donate (Mellström and Johannesson, 2008). For a more "selfish domain" it might be problematic to frame a nudge to be prosocial. For instance, save money for you own retirement in order to not burden society so others can get help.

Adding nudge technique and goal to the model

Paper 2 shows that nudge techniques and the goal of the nudge are separate components that both have an effect on the acceptability of nudges. Therefore, both nudge technique and goal of the nudge are added to the acceptance part of the model, see Figure 4. Moreover, we learned that merely framing a nudge as either pro-social or pro-self does not necessarily affect acceptance in the same way depending on the goal of the nudge. Furthermore, it is again shown that intrusiveness to freedom of choice in itself is not enough to make a nudge unacceptable, although people are more likely to find a nudge unacceptable if it also is considered to be intrusive to freedom of choice. Both nudge technique and the goal affect acceptance of nudges, but what about the stated alternative to nudge?
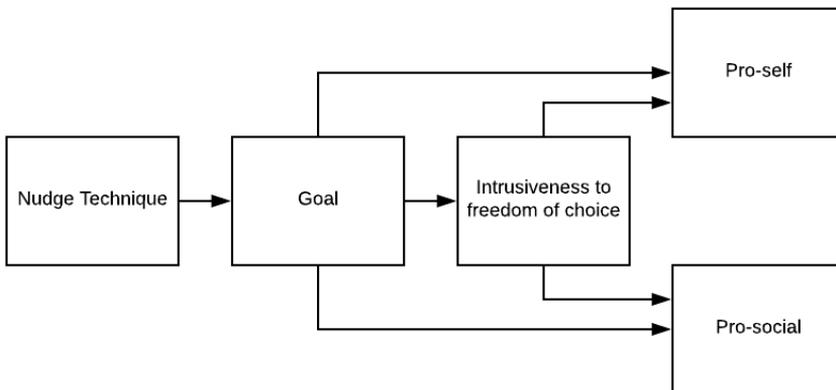
Acceptance and the alternative to the nudge

It has been previously shown that judgements and decision making are affected by reference points (e.g. Fredrick et al., 2009; Hsee and Zhang, 2010; Kahneman and Tversky, 2000; Kahneman, 2011) and the alternatives that are given to choose from (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). In light of this it possible that attitudes toward nudges also are affected by the alternative to the nudge. Davidai and Shafir (2018) showed that the attitudes towards nudges (both Type 1 and Type 2 nudges) are affected by the evaluation mode (joint or separate evaluation). The attitude towards Type 1 nudges (they used default-changing nudge techniques as the Type 1 nudge) is lower in joint evaluation with a Type 2 nudge (they used an information-providing nudge technique as the Type 2 nudge). Except when the Type 1 nudge is presented as more effective compare to the Type 2 nudge. Sunstein (2017a) found a similar result. But are nudges affected by other alternatives or opportunity costs of the nudge? To investigate this, we conducted the third study.

## *Paper 3, empirically testing the effect of alternatives*

In order to study if alternatives and opportunity cost affects the acceptance of nudges were a between-subject design web study conducted in Sweden (n = 641). The participants were recruited from a Swedish representative (online) subject pool. In order to keep things as constant as possible the same basic structure as in paper 1 and 2 was used to describe the nudge scenarios, presenting the problem followed by the nudge intervention and the intended goal of that intervention. In addition, the nudge techniques were kept constant between the contexts (a default-changing nudge technique). We constructed four nudge scenarios, two that we *a priori* categorized as pro-self (retirement savings and cancer screening) and two categorized as pro-social (organ donation and climate compensation). The participants were randomized into one of three conditions (Control, No-regulation, and Legislation-as-alternative). In the No-regulation condition it was explicitly stated (after each nudge scenario) that if the proposed default

change would not be accepted then nothing would be done at all to fix the stated problem (e.g. to get people to save more for retirement). In the legislation condition, on the other hand, it was explicitly stated (after each nudge scenario) that if the proposed default change would not be accepted then a legislation would be implemented instead to change people's behavior (e.g. change so everyone is an organ donor by law and cannot opt-out). For the control condition no consequences for rejecting the nudge were stated at all. To measure individual differences the same short version of the culture and cognition worldview scale (CCWV, Kahan, 2014) were used as in study one, while the decision style was measured with the Decision style scale (Hamilton et al., 2016)instead of the short version of the Rational-Experiential Inventory scale used in paper 1.

Surprisingly, no difference between the conditions was found for acceptance, with an acceptance rate above 60% for each condition. However, the individual differences measure of individualism (compared to communitarians) reveals that individualistic participants were differently influenced by the condition manipulations. In the control and no-regulation conditions more communitarian people were more likely to be accepting of nudges and more individualistic people were less likely to be accepting of the nudges. However, in the legislation-as-alternative condition this difference disappears. Our interpretation of this is that very communitarian individuals reject the nudge in order to get a more effective alternative in its place (i.e. the legislation alternative). We found no interaction effects with conditions for the pro-social or pro-self nudges.

Adding alternatives to the model

The alternative to a nudge did not seem to have an impact on acceptance at first glance. However, when looking at the individual differences in worldview we see that the alternatives had different effects depending on worldview. This, in addition to previous findings from Davidai and Shafir (2018) and Sunstein (2017a), who used different approaches to display alternatives compared to our study, *alternatives* is added to the acceptance part of the model in Figure 5.
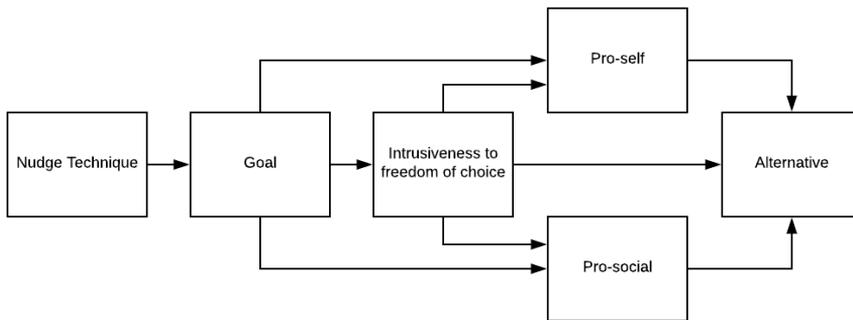
Figure 5. the nudge acceptance model expanded with the alternatives to the nudge which influence acceptance of nudges.

After paper 3 questions still remained and more questions were added. The first new question that emerged after the third study was how accepting people in Sweden are of using legislation to tackle the presented problems in study three. The lack of difference between the conditions could be explained by a high acceptance of the legislation alternatives. The other remaining question was if people are affected by the increased political polarization. So maybe who the choice architect is has an effect on acceptance of nudges?

Acceptance and the choice architect

Dolan et al. 2012 also argue that is the messenger's identity and your relation to the messenger matters a lot for the effectiveness of the nudge. In accordance with this argument, Sunstein, et al, (2018) have shown that trust in public institutions are an important factor for accepting nudges. This is also consistent with the "nudge partisan bias" which has been found by Tannenbaum et al (2017). Tannenbaum et al. found that republicans were more positive towards nudges suggested by republicans compared to the same nudge suggested by democrats (and vice versa for democrats). It is, however, unclear whether the decrease in acceptance is generated from a negative nudge partisan bias, or from the positive effect of the bias (or both). In order to test how the political affiliation with (or opposition to) the choice architect impacts the acceptance and perceived intrusiveness to freedom of choice for nudges (and whether this effect is similar for legislation) study four was conducted.

## Paper 4, empirically testing the effect of choice architect

To test how if the nudge partisan bias is generated for political parties you like or dislike and whether this effect is present in legislation as well, a web survey with 1820 participants was conducted. The study was done as a between-subject design where participants answered four scenarios, with either a nudge intervention (nudge condition) or a legislation intervention (legislation condition). Furthermore, the participants were randomized into three different political conditions where the political affiliation of the stated choice architect was manipulated. The political conditions consisted of either a left-wing coalition of parties, a right-wing coalition of parties or the Swedish democrats, resulting in a two (nudge or legislation) times three (political party) between-subject design. In order to keep a systematic approach, the same the scenarios (as either proposing nudges or legislation interventions) as in study three were used (without alternatives and modified by manipulating choice architects). After answering the scenarios each participant answered which political party in the current Swedish parliament that is furthest away from their own political views and which political party they would vote for if there was an election today. The answers from those questions were then used to measure whether each participant was in a condition where their political affiliation matched that of the choice architect (i.e. if the party they would vote for were one of the parties suggesting the intervention), if their affiliation was unmatched (i.e. the choice architect was neither a party they would vote for nor the party furthest away from them politically) or if the choice architects were their political opposition (i.e. if the party they thought was furthest away politically were one of the parties suggesting the intervention). In addition, the same individual difference measures were used as in paper 3.

The result shows a linear increase of acceptability depending on how well the political affiliation of the participants matches with the choice architect, both across the nudge conditions and the legislation condition. Furthermore, people are less likely to find the intervention (for both conditions) to be intrusive to freedom of choice the closer the participants' affiliation matches with the choice architects. Another interesting finding was that the individual difference measures had different effects depending on the political matching. For instance,

acceptance of nudges increased with age when there was a match with political affiliation and acceptance decreased with age when participants were matched with a choice architect that was furthest away from their political views. This indicates that the nudge partisan bias increases with age. Moreover, more individualistic people (compared to communitarians) were less accepting of all interventions in all conditions, except for the nudge condition when they shared the political affiliation of the choice architect.

Adding attitudes towards the choice architect to the model

The attitude towards the choice architect has a strong effect on how the nudges (and legislation) are judged. Moreover, people seem to be affected in a similar (but in opposite direction) way towards choice architects they like, and dislike compared to a more "neutral" proposer. This does not seem to be unique for nudges but is probably generalizable to all kinds of interventions. Furthermore, this seems to be true all across the political spectrum and is not a bias for the political left or the political right. Therefore, attitudes towards the choice architect are included in the model for acceptance of nudges (Figure 6).
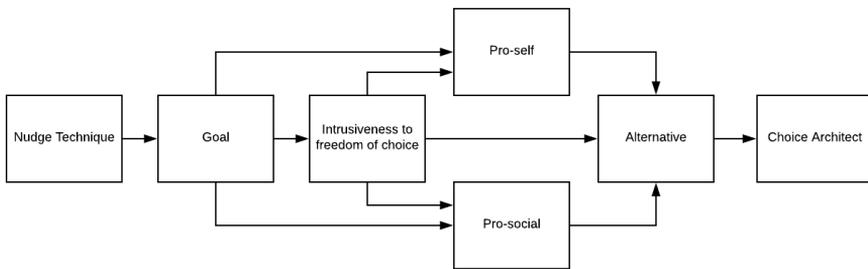
Figure 6. The nudge acceptance model expanded with attitudes towards the choice architect which influence acceptance of nudges.

The expanded nudge acceptance model

The initial model proposed in the beginning of this thesis contained a single component for acceptances; this "black box" of acceptance has been expanded to contain various components that could affect the acceptability of a nudge. These components consist of *goal of the nudge* (which targeted behaviors are increased or decreased), the *nudge technique* being used (how the intended behavior change is being targeted), who the main beneficiary is and whether the nudge is *pro-social* (is the main beneficiary society at large) and *pro-self* (is the main beneficiary the individual). Moreover, the *alternative* (which could be other forms of interventions, nudges or doing nothing) to the suggested nudge and the attitudes towards the *choice architect* (the one that proposes that the nudge should be implemented) are included since they have spillover effects towards the acceptance of nudges. Furthermore, perceived *intrusiveness* to freedom of choice is included, given that some nudges that are judged to be intrusive can still be found to be acceptable given a good enough trade-off. The final acceptance model can be seen in Figure 7.
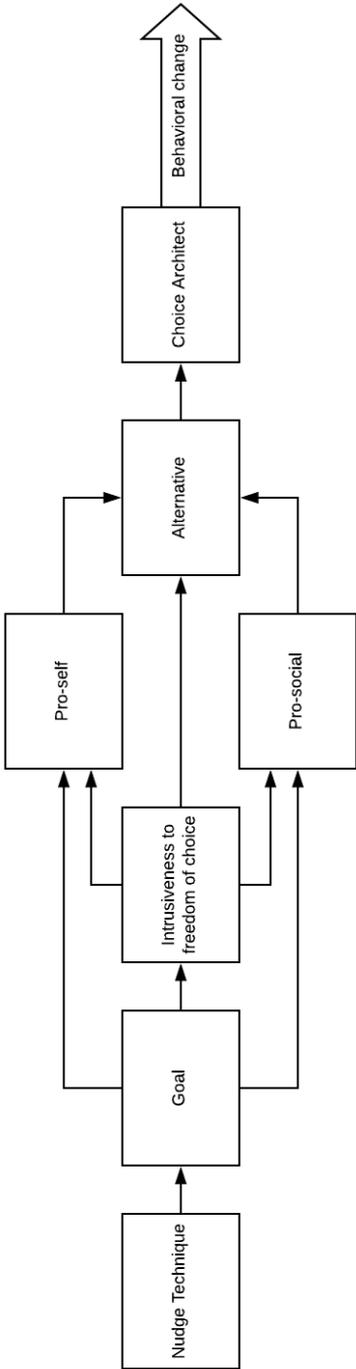
Figure 7. The final version of the nudge acceptance model with all the components that influences the acceptance of nudges.

General Discussion

In this thesis I set out to explore variables that influence acceptance of nudges. The good news for nudges is that a majority of people find most of the nudges acceptable. This finding seems to be consistent among both my own studies (paper 1-4) but also with other studies conducted since 2016 (e.g. Jung and Mellers, 2016; Reisch, Sunstein and Gwozdz, 2017; Sunstein, 2016a; Sunstein, 2017a; Sunstein, Reisch and Rauber, 2017; Loibl et al., 2018; Sunstein et al., 2018). There is, however, variance regarding when people find nudges acceptable. The main contribution of this thesis to the nudge literature is that it provides a general framework for thinking about the design of nudge techniques, but more specifically which factors that can influence the acceptability of nudges.

The starting point was to investigate whether pro-self nudges are more likely to be judged acceptable compared to pro-social nudges. In other words, who the beneficiary of the intended nudge is (and are perceived to be by the general public) and if that affects acceptance of the nudge. There is support for the idea that the nudges *a priori* categorized as pro-self in paper 1 were more accepted compared to the pro-social nudges. The same result is found in paper 3 and 4. Moreover, most of the nudges across all four studies were judged to be either equally (or more) beneficial for society at large compared to individuals. Nudges that were judged to be more beneficial for individuals (compared to society) were however the ones with the highest acceptance levels. Even if the perceived beneficiary of the nudge matters for acceptance we learned from paper 2 that simply focusing on the beneficiary by explicitly stating the intent behind a nudge as pro-social or pro-self is not enough to effect acceptance. However, there are tendencies that some domains are better than others to present as beneficial for individuals rather than beneficial for society and vice versa. Even if the evidence is indirect I intuitively think that selling a nudge as pro-self when people perceive it to be pro-social could generate a dislike since it decreases the prosociality of the behavior.

I draw two main conclusions from this. Firstly, there is evidence that pro-self nudges in fact are more acceptable compared to pro-social

nudges. This is also found by Sunstein (2017a) but he gives another explanation for the dislike of the nudge namely that people tend to dislike nudges that affect their own wallet (e.g. donating to charity set as a default). This is a typical pro-social nudge which benefits others over the individuals themselves (and the loss is salient). Secondly, people find nudges to be beneficial for both society at large and individuals. However, when there is a difference in perceived benefit then at least in my studies (and probably the nudge literature in general), people do find nudges to be more pro-social rather than pro-self. This might be problematic given both that pro-self nudges seem to be more acceptable and that the justification that nudges are libertarian paternalism rather than only paternalism (Thaler and Sunstein, 2008) is based on the pro-self versions of nudges. In other words, nudges might be more problematic than previously claimed by Sunstein (e.g. Sunstein, 2014; Sunstein, 2015; Sunstein 2017a) on a theoretical level, since many nudges that are being used are more pro-social rather than pro-self. However, this might not be as problematic as one might assume for policymakers, given that the acceptance for nudges tends to be high even for pro-social nudges. The structure for the scenarios introduced in paper 1; (1) to present the problem, (2) to present the nudge and (3) to present the intended goal with the nudge, are an important way to think about nudges for policymakers and academics alike. It is far too easy to lump these three factors together in a single unit which can lead to false conclusions when findings are generalized across contexts. For instance, an example of a false conclusion is to mistake the dislike of the goal of the nudge for the dislike of the nudge itself. Furthermore, Paper 1 contributes to the nudge literature by empirically showing that acceptance of nudges is affected by who the beneficiary is judged to be (society at large and or individuals) and that individual differences such as worldview influence acceptance while education does not necessarily affect acceptance. This conclusion was strengthened with the results of paper 3 and paper 4.

The findings in Paper 2 support and further expand the notion that the goal should be separated from the nudge technique. The information-providing techniques were the most acceptable and also had the most stable result across contexts. The information-providing nudge techniques from paper 1 (i.e. Avoiding tax evasion and Food labeling) also had high acceptability (80% or above). The variance of acceptance is higher for the default-changing nudges and the emotion-

eliciting nudge techniques. Even if the different default-changing nudges are conceptually similar they vary in how the default (change) is implemented. This makes it hard to do a perfect comparison between the same kinds of nudge techniques. However, regardless of whether the difference in acceptance within domains found in paper 2 is driven by the exact implementation of emotion-eliciting or default-changing techniques this still shows that the acceptability of a nudge is affected both by the nudge techniques being used and by the goal the nudge is trying to achieve (e.g. decrease smoking). However, since the acceptance for the same kind of default-changing nudge techniques (changing opt-in to opt-out) varies in paper 3 and 4 strengthens the conclusion that goals, and nudge techniques play a role in influencing acceptance of nudges. Moreover, we learned from paper 2 that some nudge techniques are more likely to be acceptable than others when they are judged to be intrusive to freedom of choice. For instance, default-changing nudge techniques were more likely to be acceptable compared to emotion-eliciting techniques when they were found to be intrusive, which shows that there is something more than the goal and intrusiveness that factors in while people judge nudges. Even though nudges that are judged to be intrusive to freedom of choice are more likely to not be acceptable, I find that some level of intrusiveness to freedom of choice seems to be acceptable given that the goal or the benefits from the nudges is judged good enough. As counterintuitive as it might sound, that could be why nudges have been found to be more acceptable when they are more effective (e.g. Petrescu et al., 2016; Sunstein 2017a; Davidai and Shafir, 2018). This effect should, however, have a threshold limit. An unavoidable nudge would most likely be seen as too intrusive (unless the goal is undeniably good).

A policymaker (or some other choice architect) can use various alternatives to promote behavioral change, and the way these alternatives are being presented to the general public could potentially matter. The acceptance comparison in our studies was done in a separate evaluation and not in a joint evaluation with other alternatives. It has previously been shown that the evaluation mode can matter for judgements in general ((Bazerman et al., 1992; Hsee et al, 1999; Hsee and Zhang, 2010) and for attitudes for different nudge techniques (Davidai and Shafir, 2018). This leads to another important question to explore: is acceptability of nudges influenced by a more paternalistic alternative (e.g. legislation)? This was tested in both joint (paper 3) and

separate (paper 4) evaluation mode. Surprisingly enough we did not find any real difference for acceptance of nudges between the evaluation mode (legislation was a bit lower in joint evaluation as a consequence of high acceptance for nudges). Adding a less paternalistic alternative (no intervention at all) to the nudge in joint evaluation did not change the acceptance compare to single evaluation. Although paper 3 is a first study only (and on a Swedish population) this finding suggests that policymakers do not risk decreasing the appeal of the (less paternalistic) nudge when presenting them jointly with a more effective alternative (and more paternalistic).

However, nudges are not always more acceptable than legislation. In paper 4 we found that people can find legislations more acceptable compared to nudges, when the legislation is suggested by a choice architect with matching political viewpoint as the person judging the interventions and the nudge is suggested by a choice architect that is from the opposite political spectrum. This result was found for people from both the left and the right side on the political spectrum and shows that the judgment of the intervention is very likely to be influenced by the attitudes towards the choice architect (both negatively and positively). In order words, the political affiliation of the choice architect has a big influence on acceptance of nudges. This symmetric partisan-bias is present both for nudges and for legislation which is the reason people can rate legislation from the right party as better than a nudge from the wrong party. It is however an open question whether people accept the intervention (or reject the opposition's intervention) because their judgements are made with different standards or if they are sacrificing the alternative in order to signal their support (or opposition) for the choice architect. Krijnen et al., (2017) framework of choice architecture 2.0 suggests that it might be sensemaking that drives people to signal their motivation by supporting (or opposing) the intervention from the choice architect. The other possible explanation is that people engage in motivated reasoning (Kunda, 1990) and their judgment of the intervention shifts accordingly. It might also be a combination of both motivated reasoning and social motivation that drives the effect, or different mechanisms for different people. This shows that it is important to measure acceptance of nudges prior to implementing them in order to avoid that they backfire and potentially a cause negative attitude towards the choice architect.

Outside the scope of the expanded model for behavior change through nudges, are individual differences amongst people. Although these differences affect both which nudge technique is likely to work and when a nudge is deemed acceptable, individual differences are not included in the model since the model focuses on the nudge and its components and not the individual being nudged. With that said it is important to have knowledge of the group that is being targeted by the nudge. I have shown in my studies that individualists (compared to communitarians) are more likely to reject nudges in all conditions except when the alternative to the nudge is more paternalistic or when the nudge is proposed by their own political party. The negative effect of individualism on acceptance have been replicated by Jung and Mellers (2016) who also showed that more reactant people oppose nudges they view them as autonomy-threatening and people that were more empathetic were more prone to support nudges in general. Several studies have shown that the political orientation affects the acceptance of nudges (e.g. Jung and Mellers, 2016; Sunstein 2017a; Sunstein, Reisch and Rauber, 2017). However, this might not be explicitly tied to their political opinion but rather to the matching of political viewpoints with the choice architect. Decision styles (if people are prone to intuitive or reflective thinking) or even education level seems to affect the judgement but to a lesser degree than worldview or attitudes towards the choice architect. There are most certainly many more individual differences that influence acceptance of nudges, but it is quite remarkable that one of the biggest factors (that I have found) seems directed towards the choice architect rather than how the actual nudge is designed.

Limitations

The acceptance part of the model I have suggested is based on the results in the studies conducted during this thesis work and as always in research there are caveats and limitations. The measurements of acceptance used in these studies might be one of these. There are several ways of measuring attitudes towards nudges and different measurements have been used by different researchers, either single questions (e.g. my papers, Sunstein, 2017a) or an index of several attitude measures (e.g. Tannenbaum et al, 2017). However, I believe that the concept of acceptance ranges from bare minimum of support

(or even indifference) to strong support and therefore it is a good measurement for the least support needed for policies to be implementable. Moreover, I measure judged intrusiveness to freedom of choice and benefit. These constructs are naturally closely related to acceptance. However, one of the objectives with this thesis was to see whether felt intrusiveness to freedom of choice could be equated with non-acceptance, and they correlate strongly but are not interchangeable. Benefit (especially benefit for the individual) also goes hand in hand with acceptance, but even when the rated benefit is low some nudges are accepted, which suggests that benefit and non-intrusiveness are both part of what makes the nudge acceptable.

Another possible problem is how the rated acceptance translates into actual behavior. First, it is not entirely certain that people that find described nudges to be acceptable find the same nudges acceptable after they have actually been nudged, in the same way that people that oppose some nudges actually might accept them after they have been targeted by the nudge. This is something that needs further testing. However, people that do not accept a nudge that they later come across will likely overreact (in accordance with reactance theory), which could cause both problematic behavior and spillover effects towards the choice architects, which in turn would decrease their chance of implementing any new interventions (in similarity of the finding in paper 4). Furthermore, if people are not informed about nudges in advance (and maybe given a chance to object) they will probably view them as coercive manipulation instead of optional decision support. I assume this effect will be stronger for Type 1 non-transparent biasing nudges that are good for society but not the individual (at least in the eyes of the individual). It is also unclear what happens with the acceptance of nudges when people are being informed about the nudge after they have been nudged. I have preliminary evidence of this in a study of retirement choices (Hagman in preparation). In that paper, we measure acceptance for framing retirement information in line with individuals' own preferences for time or money. The group that rated acceptance without having been subjected to the nudge intervention rated the nudge as more acceptable, less intrusive to freedom of choice and more beneficial compared to the group which rated the nudge after they had been exposed to it. This was however done with a hypothetical retirement choice, so the nudge did not change any real behavior.

Future directions for nudges

Nudges in one way or the other are here to stay, they have been around for a long time in various forms and will continue to be. The question, however, is how nudges will remain as a governmental tool with the aim of generating behavioral change for citizens. Should we give up the original definition of a nudge and just use behavioral intervention for the greater good of society while not infringing too much on freedom of choice?

The (un)stability of people's preferences generates a paradox for nudges. Nudges are supposed to help individuals to engage in a behavior that is preferred by them as judged by themselves. Given that we had complete and stable preferences, then nudges would always be acceptable to use, since they simply would not work (since a homo economicus is not affected by irrational factors). On the other hand, if humans had completely unstable preferences (e.g. preferences are always generated on the spot or at least possible to completely change with choice architecture) and nudges were extremely effective, then nudges would not be acceptable at all since they change the individual's preference and therefore are coercive. Alternatively, nudges would always be acceptable since the individual's preferences would always be in line with the nudge, since the preferences would change as judged by the individual. However, it is important to be able to differentiate between preferences people would like to have but are not ready to sacrifice anything to get and goals they actually are trying to achieve. For instance, if someone prefers to be able to save up a million dollars but is not willing to forgo the needed luxuries to achieve that goal and a choice architect hears about this preference, then the choice architect could design a very efficient nudge that helps this person save every penny they possibly could. That individual would dislike his life but would still be nudged in line with his preferences. The million-dollar-saving nudge would actually be working against his preferred way of life while a choice architect could claim that the nudge is just in line with his preferences (which would be true, but not really relevant). Would this be nudging along someone's preference or not? How can/should a choice architect rank an individual's preferences? Sometimes people want the goal without walking the path and a nudge that gets them on the path could be

considered manipulative even if the nudge help people to achieve a stated goal[5].

People tend to dislike cognitive dissonance (Festinger, 1962). Cognitive dissonance is when an individual hold two contradict beliefs or values at the same time, which induce a mental discomfort. This can arise when people are engaging in a behavior that is contradicting to their preferences. Nudges that help individuals to have a consistent behavior might be acceptable since this would resemble their true (or stable) preference, and therefore decreases the cognitive dissonance for the individual. For example, people that feel like it is important to them to be environmentally friendly could be helped to engage in (or even helped to desire) what is judged to be the most environmentally friendly options in each choice situation. This applies especially if it decreases the need for the individual to keep track of new knowledge regarding, for example, environmental consequences of behaviors or products. This is consistent with Sunstein's (2017a) argument that people like Type 1 nudges that make it easier for people to make decisions they find burdening and nudges can therefore increase people's power of agency.

There is however a slippery slope when nudges start to depart from the initial idea that the nudge should be good for the person being nudged and start to promote behavior for the greater good. Another take on when nudges are unacceptable is Thaler's idea of sludges (2018). A "sludge" is similar to nudges in the way they are supposed to affect behavior (i.e. the design or the techniques used are the same), with one important exception, the goal of the sludge. A sludge is according to Thaler designed to either discourage behavior that is in line with the individual's best interest or encourage self-defeating behaviors. It is a bit unclear where the line between a sludge and regular marketing is drawn (which is also problematic for nudges) and sometimes the difference between a nudge, a sludge and bad design might simply be the moral reasoning behind them. For instance, if I design a store without a second thought (or even by random choice) that is not really nudging or sludging since there is a lack of intent behind it. If my intent behind the store is making people buy as many

---

[5]The origin of the wanting the end without wanting the means to achieve it idea, is credited to a discussion with Luc Bovens and Andreas Schmidt on a conference in 2017.

of my (unhealthy) products as possible and maybe even offer easy store-credit that could be judged to be a sludge. However, if the products the store is promoting with behavioral design are considered healthy the same design might be considered a nudge. One question is whether my "nudge-store" can turn into a "sludge-store" depending on the general view of what is considered (un)healthy. Moreover, I think both Thaler and Sunstein would agree that tricking people into overpaying for productions with anchoring effects or framing a failure to buy as loss for the shoppers would fall into the sludge category. But what if that profit would be donated as climate compensation for the store? Would that turn the sludge it into a nudge? Would it be more of a nudge and less of a sludge if the costumers were told about it or not? Furthermore, what happens when the individual has "bad preferences"? For instance, if the consensus in a company (and among the staff) is that men and women are better at different kinds of work assignments, would it then be considered nudging to design the hiring process to increase gender inequality between departments? Since some behaviors (or preferences) are considered wrong by society in general (which like individuals' preferences is also changing over time) is it hard to say that the acceptance of a nudge only depends on each individuals' preferences. This leads to the conclusion that a nudge could either only be a concept of behavioral change (neither good nor bad; a nudge and a sludge are then the same thing with different goals) or considered a tool to increase the greater good for individuals (making nudges virtually impossible to design for more than one person at the time). This is similar to what Hansen (2016) calls the technical definition of nudging. Hansen argues that in a strict technical sense it should be possible to nudge "for good", "for bad" or for profit, since the nudge is about how people are affected and not the motive behind the behavioral change. Guala and Mittones (2015) idea of abandoning the welfarism idea of nudging, at least in a strict sense that a nudge needs to make each individual better off according to their own preferences, might actually increase the acceptance for nudges in the general public. It should be easier for policymakers to implement a nudge while proclaiming that the intent behind the nudge is to increase (or decrease) a behavior that is considered better for society (e.g. to decrease the ecological footprint) rather than that the purpose of the nudge is to help individuals act in accordance with what their own preferences should be.

Using nudges to increase welfare for society at large is still a good goal. Even though it might be hard to defend nudges on the same theoretical grounds when it departs from the principle of always being in line with the individuals' own preferences. Nudges are still easier to change compared with more paternalistic approaches (e.g. legislation), regardless of whether the change is because of a faulty design, unintended consequences or a change in the view of what a good behavior is. Designing nudges and targeting behaviors that are beneficial both for individuals and for society at large is the most acceptable (and most likely effective) way of using nudges.

Concluding remarks

Acceptance of nudges is important for policymakers to get an efficient nudge, but also in order to avoid bad will from the public. In this thesis I have presented a model for creating an acceptable nudge. I have argued and empirically shown that is it important to think about what you are trying to achieve with the nudge (the goal), how you try to achieve it (the nudge technique being used), and who benefits from it (individuals, society or both). More pro-self nudges tend to be more acceptable than pro-social nudges. Furthermore, presenting alternatives with more paternalistic interventions (at least alongside default-changing nudges) does not seem to decrease acceptance of the nudge. Moreover, it is important to know your targeted population since individual differences might affect acceptance, especially attitudes towards the choice architect themselves. If the policymaker does not have political support, it will (probably) be hard to implement a nudge but still easier compared to legislation. Taken together, the findings suggest that there is more to creating an acceptable nudge than to merely take a nudge technique that was acceptable in one context and apply it in another. However, even though nudges are considered to be more beneficial to society at large compared to individuals, they are still accepted in a majority of tested cases and therefore still a good tool for policymakers to use regardless if they are considered more pro-social compared to pro-self.

References

Baldwin, R. (2014). From regulation to behaviour change: Giving Nudge the
   third degree. *The Modern Law Review, 77(6)*, 831-857.

Bard, J. S. (2012). Lack of political will and public trust dooms presumed
   consent. *The American Journal of Bioethics, 12(2)*, 44-46.

Barton, A., & Grüne-Yanoff, T. (2015). From libertarian paternalism to
   nudging—and beyond. *Review of Philosophy and psychology, 6(3)*, 341-
   359.

Bazerman, M. H., Loewenstein, G. F., & White, S. B. (1992). Reversals of
   preference in allocation decisions: Judging an alternative versus choosing
   among alternatives. *Administrative Science Quarterly*, 220-240.

Bellé, N., Cantarelli, P., & Belardinelli, P. (2018) Prospect Theory Goes
   Public: Experimental Evidence on Cognitive Biases in Public Policy and
   Management Decisions. *Public Administration Review.*

Bovens, L. (2009). The ethics of Nudge. *In Preference change: Approaches
   from philosophy, economics and psychology (Vol. 42).* Springer Science
   & Business Media.

Brehm, S. S., & Brehm, J. W. (2013). *Psychological reactance: A theory of
   freedom and control.* Academic Press.

Brighton, H., & Gigerenzer, G. (2008). Bayesian brains and cognitive
   mechanisms: Harmony or dissonance. *The probabilistic mind: Prospects
   for Bayesian cognitive science, ed. N. Chater & M. Oaksford, 189-208.*

Cadena, X., & Schoar, A. (2011). Remembering to Pay? Reminders vs.
   Financial Incentives for Loan Payments (No. 17020). National Bureau of
   Economic Research, Inc.

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and
   regulate thought. *Trends in Cognitive Sciences, 9,* 393-399.

Costa, D. L., & Kahn, M. E. (2013). Energy conservation "nudges" and
   environmentalist ideology: Evidence from a randomized residential
   electricity field experiment. *Journal of the European Economic
   Association, 11(3)*, 680-702.

Dadvidai, S., & Shafir, E. (2018). Are 'nudges' getting a fair shot? Joint
   versus separate evaluation. *Behavioural Public Policy,* 1-19.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user
   acceptance of information technology, *MIS Quarterly, 13 (3): 319–340,*
   doi:10.2307/249008

Davis, F. D.; Bagozzi, R. P.; Warshaw, P. R. (1989), User acceptance of
   computer technology: A comparison of two theoretical models,
   Management Science, 35: 982–1003, doi:10.1287/mnsc.35.8.982

de Wijk, R. A., Maaskant, A. J., Polet, I. A., Holthuysen, N. T., van Kleef, E., & Vingerhoeds, M. H. (2016). An in-store experiment on the effect of accessibility on sales of wholegrain and white bread in supermarkets. *PloS one, 11(3),* e0151915.

Dickert, S., Västfjäll, D., Mauro, R., & Slovic, P. (2015). The feeling of risk: Implications for risk perception and communication. *In H. Cho, T. Reimer, & K. A. McComas (Eds.), The SAGE handbook of risk communication (pp. 41–54).* Thousand Oaks, CA: Sage Publications

Dickert, S., Västfjäll, D., & Slovic, P. (2015). Neuroeconomics and dual information processes underlying charitable giving. *In E. A. Wilhelms & V. F. Reyna (Eds.), Neuroeconomics, judgment, and decision making (pp. 181–199).* New York: Psychology Press.

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.

Domínguez, J., & Rojas, J. L. (2013). Presumed consent legislation failed to improve organ donation in Chile. *In Transplantation proceedings* 45(4), pp. 1316-1317. Elsevier.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences, 7(10),* 454-459.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science, 8(3)*, 223-241.

Fechner, G. T. (1860). *Elemente der Psychophysik* Elements of psychophysics]. Leipzig: Breitkopf & Härtel.

Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: public attitudes towards overt and covert Nudges. *Judgment and Decision Making, 8(3),* 202-213.

Felsen, G., & Reiner, P. B. (2015). What can neuroscience contribute to the debate over nudging?. *Review of Philosophy and Psychology, 6(3),* 469-479.

Festinger, L. (1962). Cognitive dissonance. *Scientific American, 207(4),* 93-107

Frederick, S., Novemsky, N., Wang, J., Dhar, R., & Nowlis, S. (2009). Opportunity cost neglect. *Journal of Consumer Research, 36(4)*, 553-561.

Fudenberg, D., & Levine, D. K. (2006). A dual-self model of impulse control. *American economic review, 96(5),* 1449-1476.

Furth-Matzkin, M., & Sunstein, C. R. (2017). Social Influences on Policy Preferences: Conformity and Reactance. *Minnesota Law Review., 102*, 1339.

Gigerenzer, G. (2014). *Risk savvy: how to make good decisions*. New York: Viking

Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of philosophy and psychology, 6(3),* 361-383.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. *In Simple heuristics that make us smart (pp. 3-34).* Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* Cambridge university press.

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies, 29(1),* 1-17.

Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics & Philosophy*, *32*(3), 463-483.

Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory?. *Minds and Machines, 26(1-2),* 149-183.

Goodwin, T. (2012). Why we should reject 'nudge'. *Politics, 32(2),* 85-92.

Guala, F., & Mittone, L. (2015). A political justification of nudging. *Review of Philosophy and Psychology, 6(3),* 385-395.

Hagman, W & Västfjäll, D. (in preparation). Preference for time or money: Acceptance of nudging the retirement choice.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics, 148,* 14-31.

Hamilton, K., Shih, S. I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment, 98(5),* 523-535.

Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove?. *European Journal of Risk Regulation, 7(1)*, 155-174.

Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation, 4(1)*, 3-28.

Hardin, G. (1968). The tragedy of the commons. *Science, 162(3859),* 1243-1248.

Haselager, P., van Dijk, J., & van Rooij, I. (2008). A lazy brain? Embodied embedded cognition and cognitive neuroscience. *Handbook of cognitive science: An embodied approach, 5,* 273-287.

Haselton, M. G.; Nettle, D. & Andrews, P. W. (2005). The evolution of cognitive bias *(PDF). In D. M. Buss (Ed.), The Handbook of*

Evolutionary Psychology: Hoboken, NJ, US: John Wiley & Sons Inc. pp. 724–746.

Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy, 18(1)*, 123-136.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science, 12(6)*, 973-986.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychological bulletin, 125(5)*, 576.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science, 5(4)*, 343-355.

Huntsinger, J. R., & Schnall, S. (2013). Emotion-Cognition Interactions. *In D. Riesberg (Ed.), Oxford Handbook of Cognitive Psychology (pp. 571-584).* Oxford University Press: New York, NY.

Huntsinger, J. R., Isbell, L. M., & Clore, G. L. (2014). The affective control of thought: Malleable, not fixed. *Psychological review, 121(4)*, 600.

Ly, K., Mazar, N., Zhao, M., Soman, D. (2013). A practitioner's guide to nudging. *Research Report Series Behavioural Economics in Action*. Retrieved from Rotman School of Management, University of Toronto.

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives?. *Science, 302(5649)*, 1338-1340.

Jung, J. Y. & Mellers, B. A. (2016). American attitudes toward nudges. *Judgment and Decision Making, 11(1)*, 62–74

Kahan, D.M. (2014). Cultural cognition as a conception of the cultural theory of risk. *In Handbook of risk theory, ed. S. Roeser.* Dordrecht: Springer.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist, 58(9)*, 697.

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

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

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American psychologist, 64(6),* 515.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47(2),* 263-292.

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values and frames.* New York: Cambridge University Press and the Russell Sage Foundation.

Kallbekken, S., & Sælen, H. (2013). 'Nudging' hotel guests to reduce food waste as a win–win environmental measure. *Economics Letters, 119(3)*, 325-327.

Kapsner, A., & Sandfuchs, B. (2015). Nudging as a threat to privacy. *Review of Philosophy and Psychology, 6(3)*, 455-468.

Klein, G. (1999). *Sources of Power: How People Make Decisions (Vol. 1)*. The MIT Press.

Kirchler, M., Andersson, D., Bonn, C., Johannesson, M., Sørensen, E., Stefan, M., Tinghög, G., & Västfjäll, D. (2017). The effect of fast and slow decisions on risk taking. *Journal of Risk and Uncertainty*. DOI: 10.1007/s11166-017-9252-4

Krijnen, J. M., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy, 3(2)*, i-18.

Kosters, M., & Van der Heijden, J. (2015). From mechanism to virtue: Evaluating Nudge theory. *Evaluation, 21(3)*, 276-291.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin, 108(3)*, 480.

Lapan, S. D., Quartaroli, M. T., & Riemer, F. J. (Eds.). (2011). *Qualitative research: An introduction to methods and designs* (Vol. 37). John Wiley & Sons.

Lepenies, R., & Małecka, M. (2015). The institutional consequences of nudging–nudges, politics, and the law. *Review of Philosophy and Psychology, 6(3)*, 427-437.

Lehner, M., Mont, O., & Heiskanen, E. (2016). Nudging–A promising tool for sustainable consumption behaviour? *Journal of Cleaner Production, 134,* 166-177.

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge University Press.

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.

Loibl, C., Sunstein, C. R., Rauber, J., & Reisch, L. A. (2018). Which Europeans like nudges? Approval and controversy in four European countries. *Journal of Consumer Affairs.* DOI: 10.1111/joca.12181

Mellström, C., & Johannesson, M. (2008). Crowding out in blood donation: was Titmuss right?. *Journal of the European Economic Association, 6(4)*, 845-863.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza

vaccination rates. *Proceedings of the National Academy of Sciences, 108(26)*, 10415-10420.

Mills, C. (2015). The heteronomy of choice architecture. *Review of Philosophy and Psychology, 6(3)*, 495-509.

Mols, F., Haslam, S. A., Jetten, J., & Steffens, N. K. (2015). Why a nudge is not enough: A social identity critique of governance by stealth. *European Journal of Political Research, 54(1)*, 81-98.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological review, 115(2)*, 502.

Nagatsu, M. (2015). Social nudges: their mechanisms and justification. *Review of Philosophy and Psychology, 6(3)*, 481-494.

Norris, P., Pacini, R., & Epstein, S. (1998). *The Rational-Experiential Inventory, short form.* University if Massachusetts at Amherst.

O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review, 89(1)*, 103-124.

Persson, E., Asutay, E., Hagman, W., Västfjäll, D., & Tinhög, G. (2018). Affective response predicts risky choice for fast, but not slow, decisions. *Journal of Neuroscience, Psychology, and Economics.*

Petrescu, D. C., Hollands, G. J., Couturier, D. L., Ng, Y. L., & Marteau, T. M. (2016). Public acceptability in the UK and USA of nudging to reduce obesity: the example of reducing sugar-sweetened beverages consumption. *PLoS One, 11(6)*, e0155995.

Rains, S. A. (2013). The nature of psychological reactance revisited: A meta-analytic review. *Human Communication Research, 39(1)*, 47-73.

Reisch, L. A., Sunstein, C. R., & Gwozdz, W. (2017). Beyond carrots and sticks: Europeans support health nudges. Food Policy, 69, 1-10.Sunstein, C. R. (2016) 'Do people like nudges?' *Administrative Law Review 68:* 177-210.

Ritov, I., & Baron, J. (1992). Status-quo and omission biases. *Journal of risk and uncertainty, 5(1)*, 49-61.

Rivers, N., Shenstone-Harris, S., & Young, N. (2017). Using nudges to reduce waste? The case of Toronto's plastic bag levy. *Journal of environmental management, 188*, 153-162.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica, 5(17)*, 61-71.

Schmidt, A. T. (2017). The Power to Nudge. *American Political Science Review, 111(2)*, 404-417. https://doi.org/10.1017/S0003055417000028

Schubert, C. (2017). Green nudges: Do they work? Are they ethical?. *Ecological economics*, *132*, 329-342.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science, 18(5)*, 429-434.

Schwartz, J., Riis, J., Elbel, B., & Ariely, D. (2012). Inviting consumers to downsize fast-food portions significantly reduces calorie consumption. *Health Affairs, 31,* 399–407.

Sharot, T. (2011). The optimism bias. *Current biology, 21(23)*, R941-R945.

Sheffer, L., Loewen, P. J., Soroka, S., Walgrave, S., & Sheafer, T. (2018). Nonrepresentative representatives: an experimental study of the decision making of elected politicians. *American Political Science Review, 112(2)*, 302-321.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics, 69(1)*, 99-118.

Slovic, P., & Västfjäll, D. (2010). Affect, moral intuition, and risk. *Psychological Inquiry, 21,* 387–398.

Strömbäck, C., Lind, T., Skagerlund, K., Västfjäll, D., & Tinghög, G. (2017). Does self-control predict financial behavior and financial well-being? *Journal of Behavioral and Experimental Finance, 14,* 30–38. doi: 10.1016/j.jbef.2017.04.002

Sunstein, C. R. (2014). *Why nudge?: The politics of libertarian paternalism*. Yale University Press.

Sunstein, C. R. (2015). *Choosing not to choose: Understanding the value of choice.* Oxford University Press.

Sunstein, C. R. (2016a) 'Do people like nudges?' *Administrative Law Review 68:* 177-210.

Sunstein, C. (2016b). Fifty Shades of Manipulation. *Journal of Marketing Behavior, 1(3-4)*, 213-244.

Sunstein, C.R. (2016c). People prefer System 2 nudges (kind of). *Duke Law Journal, 66(1)*, 121-168.

Sunstein, C. R. (2017a). *Human Agency and Behavioral Economics: Nudging Fast and Slow*. Springer.

Sunstein, C. R. (2017b). Default rules are better than active choosing (Often). *Trends in Cognitive Sciences, 21(8)*, 600-606.

Sunstein, C. R. (2017c). Nudges that fail. *Behavioural Public Policy, 1(1)*, 4-25.

Sunstein, C, R., Reisch, L, A., & Kaiser, M. (2018) Trusting nudges? Lessons from an international survey, *Journal of European Public Policy,* DOI: 10.1080/13501763.2018.1531912

Sunstein, C. R., Reisch, L. A. & Rauber, J. (2017) A worldwide consensus on nudging? Not quite, but almost. *Regulation & Governance, 12(1),* 3-22.

Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making, 31(3)*, 355-366.

Tannenbaum, D., Fox, C. R., & Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour, 1(7),* 0130.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making, 12(3)*, 183-206.

Thaler, R. H. (2018). nudge, not sludge. *Science, 361,* 431-431.

Thaler, R. H., & Shefrin, H. M. (1981). An economic theory of self-control. *Journal of political Economy, 89(2)*, 392-406.

Thaler, R., & Sunstein, C, R. (2003). "Libertarian Paternalism". *The American Economic Review 93:* 175–79.

Thaler, R., & Sunstein, C, R.. (2008). *Nudge: improving decisions about health, wealth and happiness.* New Haven: Yale University Press.

Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L., & Västfjäll, D. (2016). Intuition and moral decision-making—the effect of time pressure and cognitive load on moral judgment and altruistic behavior. *PLoS ONE, 11(10)*, e0164012. doi: 10.1371/journal.pone.0164012

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185(4157)*, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211(4481)*, 453-458.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics, 106(4)*, 1039-1061.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty, 5(4)*, 297-323.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *Management Information Systems quarterly*, 425-478.

Västfjall, D., Slovic, P., Burns, W., Erlandsson, A., Koppel, L., Asutay, E., & Tinghög, G. (2016). The arithmetic of emotion: Integration of incidental and integral affect in judgments and decisions. [Review]. *Frontiers in Psychology, 7*. doi: 10.3389/fpsyg.2016.00325

Västfjäll, D., & Slovic, P. (2013). Cognition and emotion in judgment and decision making. In M. D. Robinson, E. R. Watkins, & E. Harmon-Jones

(Eds.), *Handbook of cognition and emotion* (pp. 252–271). New York: Guilford Press.

Weber, E. H. (1834). De subtilitate tactus [On the sensitivity of the touch sense]. In E. H. Weber, *De pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae* (pp. 44–174). Leipzig: Koehler.

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. *In J. Nielsen & R. L. Mack (eds.). Usability inspection methods.* New York, NY: John Wiley.

Whitman, D. G., & Rizzo, M. J. (2015). The problematic welfare standards of behavioral paternalism. *Review of Philosophy and Psychology, 6(3)*, 409-425.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance*. Psychology Press.

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

LINKÖPING STUDIES IN BEHAVIOURAL SCIENCE

186.   ELWÉR, ÅSA. Early Predictors of Reading Comprehension Difficulties. 2014. ISBN: 978-91-7519-281-9

187.   LARSON, NICLAS. Matematikämnet och stadiebytet mellan grundskolan och gymnasieskolan. En enkät- och klassrumsstudie. 2014. ISBN: 978-91-7519-196-6

188.   LY, HOA. Use of a Smartphone Application in the Treatment of Depression – The New Wave of Digital Tools for Psychological Treatment. 2015. ISBN: 978-91-7519-136-2

189.   HANSSON, PER-OLOF. New ways of learning. Participatory action research and Kenyan runners' appropriation of smartphones to improve their daily lives and participation in m-learning. 2015. ISBN: 978-91-7519-124-9

190.   BOLLDÈN, KARIN. Online teaching practices. Sociomaterial matters in higher education settings. 2015. ISBN: 978-91-7519-123-2

191.   MALMQUIST, ANNA. Pride and Prejudice. Lesbian families in contemporary Sweden. 2015. ISBN: 978-91-7519-087-7

192.   AVBY, GUNILLA. Evidence in Practice. On Knowledge Use and Learning in Social Work. 2015. ISBN: 978-91-74519-088-4

193.   FORSBERG, CAMILLA. Students' Perspectives on Bullying. 2016. ISBN: 978-91-7685-874-5

194.   MUHRMAN, KAROLINA. Inget klöver utan matematik. En studie av matematik i yrkesutbildning och yrkesliv. 2016. ISBN: 978-91-7685-851-6

195.   SKAGERLUND, KENNY. Magnitude Processing in Developmental Dyscalculia. A Heterogeneous learning disability with different cognitive profiles. 2016. ISBN: 978-91-7685-831-8

196.   EKEBLAD, ANNIKA. A Randomized Trial of Interpersonal Psychotherapy and Cognitive Behavioral Therapy for Major Depressive Disorder – Predictors of process and outcome. 2016. ISBN: 978-91-7685-803-5

197.   MÅNSSON, KRISTOFFER N.T. Restructuring the Socially Anxious Brain. Using magnetic resonance imaging to advance our understanding of effective cognitive behaviour therapy for social anxiety disorder. 2016. ISBN: 978-91-7685-688-8

198.   SILFVERNAGEL, KRISTIN. Individually tailored internet-based cognitive behavioural therapy for adolescents, young adults and older adults with anxiety. 2017. ISBN: 978-91-7685-566-9

199.   ELOFSSON, JESSICA. Children´s early mathematics learning and development. Number game interventions and number line estimations. 2017. ISBN: 978-91-7685-517-1

200.   ÅHLFELDT, EMANUEL. Hållbart utvecklingsarbete i vård och omsorg. Ett institutionellt perspektiv på projekt i en professionell och byråkratisk kontext. 2017. ISBN: 978-91-7685-445-7

201.   VERNMARK, KRISTOFER. Therapeutic alliance and different treatment formats when delivering internet-based CBT for depression. 2017. ISBN: 978-91-7685-436-5

202.   FRANKL, MY. Psychotherapy for Substance Use Disorder – the importance of affects. 2017. ISBN: 978-91-7685-429-7

203.   ABDULLA, AFRAH. Readiness or resistance? – Newly arrived adult migrants' experiences, meaning making, and learning in Sweden. 2017. ISBN: 978-91-7685-428-0

204.   OLSSON, LINDA. "Count on me!". Mathematical development, developmental dyscalculia and computer-based intervention. 2018. ISBN: 978-91-7685-409-9

205.   COLLIANDER, HELENA, Being and Becoming a Teacher in Initial Literacy and Second Language Education for Adults. 2018. ISBN: 978-91-7685-304-7

206.   TOPOOCO, NAIRA. Blended Cognitive Behavior Therapy. Efficacy and Acceptability for Treating Depression in the Adult and Adolescent Population. 2018. ISBN: 978-91-7685-297-2

207.   SOLÍS MARCOS, IGNACIO. Challenges in Partially Automated Driving: A Human Factors Perspective. 2018. ISBN: 978-91-7685-296-5

208.   THORSTEN, ANJA. Berättelseskrivande i skolan. Att studera, beskriva och utveckla ett kunnande. 2018. ISBN: 978-91-7685-258-3

209.   ELEONOR, BREDLÖV EKNOR. Shaped for beauty. Vocational and gendered subjectivities in private education for the beauty industry.

210.   VESTERGREN, SARA. Psychological change as an outcome of participation in collective action. 2018. ISBN:978-91-7685-242-2

211.   MÖLLER, CLARA. Mentalizing. Competence and process. 2018. ISBN: 978-91-7685-189-0

212.   SÖDERBERG GIDHAGEN, YLVA. Psychological treatment of outpatients with substance use disorders in routine care – attachment style, alliance, and treatment outcome. 2018. ISBN: 978-91-7685-197-5