

**A CRITICAL EXAMINATION OF NUDGE ON THEORETICAL AND ETHICAL
GROUNDS**

YILING LIN

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

2018

Queen Mary University of London
Biological and Experimental Psychology
School of Biological and Chemical Sciences

STATEMENT OF ORIGINALITY

I, Yiling Lin, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: 20/04/2018

Details of collaboration and publications:

Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, Effectiveness, and Ethics.

Basic and Applied Social Psychology, 39(6), 293–306.

<https://doi.org/10.1080/01973533.2017.1356304>

Osman, M., Lin, Y., & Ashcroft, R. (2017). Nudging: A lesson in the theatrics of choice. *Basic and Applied Social Psychology*, 39(6), 311-316.

Lin, Y., Osman, M., Harris, A. J. L., & Read, D. (2018). Underlying Wishes and Nudged Choices. *Journal of Experimental Psychology: Applied*. Accepted.

Lin, Y., Osman, M., Wong, R. S., Harris, A. J. L., & Wong, W. H. (under review). Nudged Into Retirement.

Osman, M., Lin, Y., Gold, N. What factors persuade the public on social policy interventions? (in preparation).

ABSTRACT

Nudges are psychologically informed tools designed to promote behaviour change that improve health and wellbeing. As nudges gain global appeal through government administrations, their implementation as a tool within social policy are subject to stricter scrutiny. The need for inspection of the current underpinning theoretical framework of nudges is particularly critical given the diversity of nudge-type interventions and the mixed evidence for their effectiveness. The thesis first conceptualised nudge into Type 1 and Type 2 differed according to the amount of re-evaluation and re-structuring of knowledge representations needed to bring in alignment the evidence base and choice behaviour. Based on this conceptualisation, evidence base in the health domain were critically examined to determine which type of nudges is effective. The literature review showed that Type 2 nudges are more effective at sustaining behaviour change than Type 1 nudges. The thesis then explored the extent to which nudge undermines autonomy. From a welfarist perspective, nudges such as default rules bypass people's reflective thinking and make general presumption about people's choices which may not align with their underlying wishes. Indeed, two online questionnaires revealed that third-party judgements of choices made under pro-self default in pension enrolment and pro-social default in organ donation were less likely to represent the individual's "true" preference compared to when an active choice is made. The welfare consequence of implementing defaults were discussed in these respective contexts. From a transparency perspective, influencing people's behaviour without their awareness can be regarded as manipulation and hence a violation of autonomy. The online survey findings revealed that the public generally find interventions easier to identify, perceive them as more effective and acceptable, and more likely to align with people's values or interest when they are transparent, and have a positive rationale for how they work. In closing, the proposed theoretical framework corroborated the empirical findings which suggest that Type 1 nudges are more likely to bypass people's reflective thinking and infringe upon personal autonomy. Conversely, Type 2 nudges that work in a transparent manner are more likely to promote behavioural change in line with the chooser's higher-order desire. For policy makers, the important practical point is that Type 2 nudges are more likely to preserve autonomy in a way that will trigger sustainable behavioural change at a population level compared to Type 1 nudges.

Keywords: nudge; theory; evidence; ethics; autonomy; welfare; transparency

TABLE OF CONTENT

STATEMENT OF ORIGINALITY.....	2
ABSTRACT.....	3
LIST OF FIGURES AND TABLES	6
ACKNOWLEDGEMENTS.....	10
CHAPTER 1: General Introduction.....	12
1.1. Nudge: An Introduction.....	13
1.2. Effectiveness of Nudge.....	14
1.3. Theoretical Complexities.....	16
1.4. Nudge and Autonomy.....	17
1.4.1. The Welfarist Argument.....	19
1.4.2. The Transparency Argument.....	21
1.5. Summary.....	23
CHAPTER 2 – Nudge: Theoretical and Empirical Re-consideration.....	25
2.1. Critical issues with Dual-System theories.....	25
2.2. A Single System Account of Nudge.....	28
2.3. Empirical Evidence in Health Domain.....	30
2.4. Empirical Issues Concerning Nudge.....	36
2.5. General Theoretical Re-Considerations of the Nudge Evidence-base.....	38
2.6. Summary.....	39
CHAPTER 3: Nudge and Autonomy.....	40
3.1. The Welfarist Argument - “True” Preference in Default.....	41
3.1.1. Pro-self Default in Pension Enrolment.....	43
3.1.2. Pro-social Default in Organ Donation Registration.....	46
3.1.3. Bayesian Approach.....	49
3.1.4. Summary – Pro-social and Pro-self nudges.....	50
3.2. The Transparency Argument: Public Attitudes Towards Nudge.....	51
3.2.1. Other Predictors.....	55
3.2.2. Summary – Public Attitude Towards Nudges.....	57
CHAPTER 4: Underlying Wishes and Nudged Choices in Pro-self Context.....	58
4.1. Participants and Design.....	60
4.2. General Materials and Procedure.....	61
4.3. Results.....	64

4.4.	Discussion	69
CHAPTER 5: Underlying Wishes and Nudged Choices in Pro-social Context.....		72
5.1.	Participants and Design.....	74
5.2.	General Materials and Procedure.....	74
5.3.	Results.....	78
5.4.	Discussion	95
CHAPTER 6: Public Attitude Towards Nudge.....		98
6.1.	Participants and Design.....	99
6.2.	General Materials and Procedure.....	102
6.3.	Results.....	103
6.4.	Discussion	116
CHAPTER 7: General Discussion.....		117
7.1.	Summary of Thesis Objectives	117
7.2.	Research Objective 1: Theoretical and Empirical Re-consideration	118
7.3.	Research Objective 2: The Welfarist Argument.....	119
7.4.	Research Objective 2: The Transparency Argument.....	126
7.5.	A Lesson on Autonomy	132
7.6.	Conclusion	132
REFERENCE.....		135
APPENDIX A.....		157
APPENDIX B		187
APPENDIX C		196
APPENDIX D.....		210

LIST OF FIGURES AND TABLES

Chapter 2 Nudge: Theoretical and Empirical Re-consideration		
Table 1	A summary of the evidence on the effectiveness of Type 1 and Type 2 Nudges.	29
Chapter 4 Underlying Wishes and Nudged Choices in Pro-self Context		
Figure 1	Description of the enrolment systems in each experimental condition for both the UK sample and HK sample.	61
Figure 2	Mean perceived belief of an investor's (i.e. Mark's) true preference to register in the ISA scheme (Experiment 1a, UK sample) and in the MPF scheme (Experiment 1b, HK sample) across the three experimental conditions. Error bars are 95% CI.	64
Figure 3	Mean perceived belief of an investor's (i.e. Mark's) true preference to register in the ISA scheme (Experiment 2a, UK sample) and in the MPF scheme (Experiment 2b, HK sample) across the three experimental conditions. Error bars are 95% CI.	65
Figure 4	The likelihood estimates of participants registering in a retirement savings scheme under the same enrolment system as Mark across the three experimental conditions in Experiment 1a (UK Sample) and Experiment 1b (HK Sample). Error bars are 95% CI.	67
Chapter 5 Underlying Wishes and Nudged Choices in Pro-social Context		
Figure 5	The scenario described in each organ donation legislative systems: Default Opt-In, Default Opt-out, Mandated Choice and Mandatory Donor.	73
Figure 6	Mean perceived belief of the donor's (i.e. Mark's) true preference to donate across the four experimental conditions for Study 1, Study 2 and Study 3. Error bars are at 95% CI.	76
Figure 7	Mean perceived signal of the donor's (i.e. Mark's) intention to donate across the four experimental conditions for Study 2 and Study 3. Error bars are at 95% CI.	78

Figure 8	Mean perceived likelihood estimates of the relative's belief of the donor's (i.e. Mark's) belief of the donor's (i.e. Mark's) true preference to donate across the four experimental conditions for Study 2 and Study 3. Error bars are at 95% CI.	79
Figure 9	Perceived likelihood estimates of the relative's decision to consent across the four experimental conditions for Study 1, Study 2 and Study 3. Error bar at 95% CI.	81
Table 2	The combined general reasons in frequencies for opt-in and opt-out countries that John will consent to donate Mark's organs in each organ donation legislative system, where 'x' means this option was not displayed for that system.	83
Table 3	The combined general reasons in frequencies for opt-in and opt-out countries that John will refuse to donate Mark's organs in each organ donation legislative system, where 'x' means this option was not displayed for that system.	84
Figure 10	Mean perceived diagnosticity of people's intention to donate across the four experimental conditions for Study 1, Study 2 and Study 3. Error bar at 95% CI.	85
Table 4	Participants' knowledge of the current organ donation legislation in their country: Study 1 (US, opt-in system), Study 2 (European country, opt-in system), and Study 3 (European country, opt-out system).	88
Table 5	Summary of Multiple Regression of mean perceived belief of the donor's true preference to donate with predictor variables experiment, legislative system, age, gender, forbidden organ donation, organ donor, blood donor, known registered donor, agree with opt-out, willingness to donate loved one's organ, who decides in the event of death.	92
Table 6	Summary of Multiple Regression of mean perceived signal of the donor's intention to donate with predictor variables experiment, legislative system, age, gender, forbidden organ donation, organ donor, blood donor, known registered donor, agree with opt-out,	93

willingness to donate loved one's organ, who decides in the event of death.

Table 7	Summary of Multiple Regression of mean perceived likelihood estimates of the relative's belief of the donor's true preference to donate with predictor variables experiment, legislative system, age, gender, forbidden organ donation, organ donor, blood donor, known registered donor, agree with opt-out, willingness to donate loved one's organ, who decides in the event of death.	94
---------	---	----

Table 8	Summary of Multiple Regression of mean perceived likelihood estimates of the relative's decision to consent with predictor variables experiment, legislative system, age, gender, forbidden organ donation, organ donor, blood donor, known registered donor, agree with opt-out, willingness to donate loved one's organ, who decides in the event of death.	95
---------	---	----

Chapter 6 Public Attitude Towards Nudge

Table 9	Participants profile from Experiment 1, 2 and 3 combined.	100
Table 10	Description of Transparent and Non-Transparent Behavioural Interventions.	101
Table 11	Contingency table showing people's accuracy in identifying the transparency of an intervention across the five behavioural domains.	104
Figure 11	The main effect of context on mean ratings of ease of identification. Error bar at 99% CI.	105
Figure 12	The interaction between context and transparency of an intervention on mean ratings of ease of identification. Error bar at 99% CI.	106
Figure 13	Main effect of context on mean ratings of willingness to change behaviour. Error bar at 99% CI.	107
Figure 14	The interaction between context and transparency on the mean rating of willingness to change behaviour. Error bar at 99% CI.	107
Figure 15	The main effect of context on mean ratings of perceived effectiveness of an intervention. Error bar at 99% CI.	109

Figure 16	The interaction between context and transparency on mean ratings of perceived effectiveness of an intervention. Error bar at 99% CI.	109
Figure 17	The main effect of context on mean rating of public acceptability. Error bar at 99% CI.	110
Figure 18	The interaction between context and transparency on mean rating of public acceptability. Error bar at 99% CI.	111
Table 12	Summary of Multiple Regression of mean ease of identification with predictor variables transparency, sample, behaviour agent, rationale of intervention, age, gender, political affiliation, religiosity, education level.	112
Table 13	Summary of Multiple Regression of mean ease of identification with predictor variables transparency, sample, behaviour agent, rationale of intervention, age, gender, political affiliation, religiosity, education level.	113
Table 14	Summary of Multiple Regression of mean perceived effectiveness with predictor variables transparency, sample, behaviour agent, rationale of intervention, age, gender, political affiliation, religiosity, education level.	114
Table 15	Summary of Multiple Regression of mean public acceptability with predictor variables transparency, sample, behaviour agent, rationale of intervention, age, gender, political affiliation, religiosity, education level.	115

ACKNOWLEDGEMENTS

The completion of this research was funded by Queen Mary University of London Life Sciences Initiative Studentship (LSIPGRS).

I would like to express my sincere gratitude to my supervisor Dr Magda Osman. She has been my guiding lights throughout this long journey and the completion of this work would not have been possible without for her continuous moral support, patience, advice, and encouragement. Her enthusiasm and wealth of knowledge in human behaviour has inspired my ongoing passion for science and exploration. I am extremely thankful for this wonderful opportunity to complete my PhD under her supervision, it has truly been an honour. I would also like to thank both Dr Magda Osman and Prof. Richard Ashcroft for their insightful comments throughout the writing and editing of this thesis.

Special thanks to the amazing support from members of the Dynamic Learning and Decision Making lab: Agata Ludwiczak, Luigi Baciadonna, Omar Verduga, Caroline Spence, Milena Nikolioc and especially Jie-Yu Lv; for all the stimulating discussions, and for their insightful comments for various projects during my PhD study.

Last but not least, I would like to thank my family: my parents for their unconditional love and support throughout my life; my husband Steve for providing the “distraction” when I needed. I would like to thank my best friends Heidi Miu and Giang Tran for always being there for me through the ups and downs. Most importantly, I am grateful to Mr & Mrs Lam for the greatest gift of all: an education. They have inspired and encouraged me to be a better person, making me who I am today.

Dedicated to Mr & Mrs Lam

CHAPTER 1: General Introduction

In this chapter, the main focus will be on setting the critical themes of this thesis. The first section provides an introduction to nudge and how it arose from a line of research in behavioural economics. As the use of nudges gain global recognition through government administrations, their application is also subject to increasing scrutiny. More specifically, their implementation as a tool for social policy should be based on a sound theoretical framework with systematic evidence and ethical justification. The current evidence base on the effectiveness of nudge-type interventions is scarce with only a limited number of systematic reviews of domain-specific interventions. Overall, the mixed evidence base is due to heterogeneity in nudge intervention designs and a lack of randomised controlled trials at the population level. In part, this heterogeneity can be attributed to a lack inspection/scrutiny of the theoretical framework underpinning the mode of how nudges function to shape human behaviour. Following this, the second section outlines the first objective of this thesis to explore the theoretical complexities concerning nudges - drawing on two critical issues related to the current underlying mechanism of nudge - with the aim of proposing an alternative framework to account for the current evidence base of nudge based research in the health domain. The third section examines ethical issues of nudge utilization with regard to infringement of autonomy. The second objective of the thesis is to empirically examine the extent to which nudge undermines autonomy: 1) If nudges do not align with the chooser's higher order desire - the welfarist argument; and 2) If people cannot discern the nudge intervention and how it is used to change their behaviour - the transparency argument. On welfare grounds, I demonstrate how defaults that bypass people's reflective thinking typically impose value substitution. To explore the welfare consequence of implementing such defaults, a distinction between pro-self and pro-social nudges was emphasized due to their differing objectives. This motivated two empirical studies to examine third party judgements of perceived true preference of nudges in the context of pension enrolment and organ donation respectively. On transparency grounds, I illustrate how the publicity principle loses its ground and why we need to redefine transparency from the perspective of those being nudged; this led to the third empirical study to investigate public's judgement on transparency of interventions in order provide relevant information for consideration of ethical issues regarding autonomy.

1.1.Nudge: An Introduction

Economist Richard Thaler and legal scholar Cass Sunstein's seminal book on "*Nudge: Improving Decisions about Health, Wealth and Happiness*" outlined an alternative governance intervention that uses subtler approaches in the design of policies to influence the behaviour of citizens. The objective of nudges is therefore to steer people's choices in a welfare-promoting direction while preserving their freedom of choice. In their original definition, a nudge is defined as "*any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives.*" (Thaler & Sunstein, 2008, p6). This basically means that the choice context can be (cleverly) rearranged to increase the likelihood that people will select the option that is better for their health, wealth, and happiness. The concept of nudge has its roots in behavioural economics; a line of research that challenges neoclassical economic models of rational human behaviour which argued that humans are utility-maximizers with well-defined preferences and unlimited cognitive capacities. Despite this, evidence from lab experiments have illustrated that people are subject to bounded rationality, bounded self-interest, and bounded willpower (Jolls, Sunstein, & Thaler, 1998). Because of this, they rely on heuristics or rules of thumbs; such as availability, representativeness, and adjustment and anchoring when making judgement under uncertainties (Tversky & Kahneman, 1974). Although these heuristics are usually effective, they can lead to systematic biases such as optimism and overconfidence, loss aversion, status quo bias, framing, saliency bias, inconsistent preference over time, to name but a few (Thaler & Sunstein, 2008). These cognitive biases lead people to make suboptimal decisions, for example, consume more food when served with a larger plate, reduce their potential saving for retirement because of procrastination, or overconsume alcohol because of a tendency to follow the crowd.

Due to the extent that people are constrained by the choice environment in which they make decisions (i.e., how it is constructed and the way they are influenced by framing of choices), subtle changes to the way in which choices are presented can encourage decisions that promote their welfare. The concept of nudge is justified by the philosophical framework of '*libertarian paternalism*'. The libertarian aspect gives people the freedom to choose and to opt out of any arrangements if they want to, and the paternalistic aspect lies in the claim that it is legitimate for choice architects to influence people's behaviour in a way that "*make choosers better off, as judged by themselves*" (Thaler & Sunstein, 2008, p5). This relatively non-intrusive mode of

governance has been sought after by governments from across the globe. In 2010, the UK was the first to establish a “Nudge Unit” (later renamed “Behavioural Insights Team” or BIT for short) followed by the US. The BIT has since carried out large-scale experimental trials to inform policy making in a wide range of behavioural contexts such as organ donation, alcohol consumption, consumer empowerment, environment, and tax (Cabinet Office Behavioural Insights Team, 2011).

1.2. Effectiveness of Nudge

Since its inception, the views on the use of nudges range from those praising their potential benefits (Mills, 2013; Saghai, 2013; Sunstein, 2015b), to those that doubt how they are utilized (Goodwin, 2012; Mols, Haslam, Jetten, & Steffens, 2015; Osman, 2016; Selinger & Whyte, 2012) and whether they are effective (Bonell, McKee, Fletcher, Wilkinson, & Haines, 2011; Kusters & Van der Heijden, 2015; Marteau, Ogilvie, Roland, Suhrcke, & Kelly, 2011). As nudge becomes increasingly applied to policy making, there is a need to understand the different ways in which nudge can lead to ethical, effective, and sustainable behavioural change. To date, there is a lack of research at the population level to indicate the effectiveness of nudge at sustaining behaviour change in the long-term (Marteau et al., 2011). In a scoping review of 346 articles focusing on diet, physical activity, alcohol, and tobacco-related behaviours, G. J. Hollands et al. (2013) concluded that most studies were highly heterogeneous with respect to the populations, interventions, comparators or counterfactuals, outcomes, and moderators that were assessed. There were a small number of non-systematic reviews and broad overviews attempting to summarise parts of the overall evidence base, but only a few were of high-quality. Nørnberg, Houlby, Skov, & Pérez-Cueto (2016) reviewed 12 studies using interventions to promote vegetable consumption among adolescents and found no conclusive evidence for their effectiveness as the majority of studies were of weak or moderate quality. Similarly, Skov, Lourenço, Hansen, Mikkelsen, & Schofield (2013) found that a majority of 12 studies on changing eating behaviour in self-service eating settings were of very weak quality due to small sample sizes, inaccurate outcome measurement tools, short implementation durations, and a lack of descriptive information. Most recently, Lin, Osman, & Ashcroft (2017) examined a range of nudges implemented in the health domain. They found that the empirical support for nudge was not robust enough to make firm conclusions as to their current effectiveness, as too few studies make direct comparison of different nudge techniques against more conventional ways of inducing behavioural changes (e.g., taxes, fines, financial incentives) and against

control baseline conditions. Some studies were also conducted in the field, as with many field based studies the difficulties are running studies on a large sample with proper controls. Further, it is rare to find field studies that also carry out follow-ups to examine the effects of the nudges in the long term. Although the BIT has pioneered the use of randomised controlled trials to determine effectiveness of interventions with some considerable success, the Science and Technology Committee (2014) raised concerns regarding the availability of robust evaluation of data in the public domain to substantiate such claims about the success of some of its interventions.

Another issue of concern in assessing efficacy of such interventions is the context-specific nature and generalisability of these interventions. That is, what works in one context with one groups of people may have limited effect in a different setting and a different group of people. The mixed evidence-base is due to their heterogeneous design which can be attributed to a lack inspection/scrutiny of the theoretical framework underpinning the mode of how nudges function to shape human behaviour. That is, if the theoretical framework on which nudge is proposed to work is founded problematic, then this has implication for the evaluation of nudge effectiveness. In order to establish effectiveness, it becomes apparent that there is a need for a conceptually sound framework. Currently, the broad definition and types of nudges mean there is no clear consensus on what counts as nudges and how they influence people's behaviour, particularly as Thaler and Sunstein themselves offered examples of nudges that failed to conform to their initial definition (Selinger & White, 2012). For example, they offered an example of nudge in which teenagers were paid to not get pregnant while the definition of nudge explicitly stated that it should not significantly change people's incentives. As of yet there is no precise, operational definition of nudge to indicate the mechanism by which choice architecture can be altered to change behaviour (Hollands, Shemilt, et al., 2013). This becomes a particular problem if policy maker were to design and implement nudges because there isn't a clear picture of what nudge is, how it operates to change behaviour and the characteristic of the choice environment (Osman, 2016). With continuous policy interests worldwide, it is essential to establish clarity in the nudge concept so as to determine the effectiveness of interventions and to design effective interventions. The following section will outline the current underlying mechanism by which nudge is proposed to work and how this framework is potentially flawed for explaining the evidence base for nudge.

1.3.Theoretical Complexities

The proposed underlying mechanism of nudge is the dual system theory (DST) which outlines two systems working together to influence choices and behaviours. System 1 is driven by immediate feelings and triggered by the environment. It is automatic, rapid, and instinctive, requiring no cognitive engagement. System 2 is a reflective, goal-oriented system that is deliberate, self-conscious, and has limited cognitive capacity. Nudge theorists (Thaler & Sunstein, 2008) claim that the basis of poor lifestyle choices is commonly through activation of System 1 type processes. Hence, a practical way of generating positive behavioural change is to target System 1 processes by re-orienting the features of the choice contexts on which heuristics and biases are invoked; this is typically achieved covertly without the decision-maker's awareness. The distinction between these two kinds of thinking have been extensively debated in the areas of reasoning, judgment and decision making, and social cognition (Evans & Over, 1996; Sloman, 1996; Stanovich & West, 2000). Some claim that the systems are interactive (Sloman, 1996); that they operate in parallel (Epstein, 1973; Evans & Over, 1996); that they have a sequential relationship (Gilbert, 1989) whilst others claim that System 1 is the default and tends to dominate over other systems (Evans & Stanovich, 2013). Although the DST had enjoyed considerable popularity over the last few decades, it has been challenged on the basis of empirical evidence (Osman, 2004, 2013, 2014a; Osman & Stavy, 2006). The core issues are associated with the distinction between implicit and explicit processes whereby implicit processes occur without awareness, and explicit processing is deliberate and accompanied by awareness (Evans & Over, 1996; Stanovich & West, 2000). More specifically, there are two core issues with the System 1/System 2 distinction: 1) the nature of the interaction between the two systems; and 2) the lack of precision around the details of key distinctions between the two systems (Gigerenzer & Regier, 1996; Osman, 2016).

At present, the exact relationship between these two systems is unclear, and more profoundly still, whether the distinctions between Systems 1 and 2 are an over simplification of a single system (Osman, 2013; Osman, 2014). The key point being that nudge is an indirect way of influencing lifestyle choices by tapping into System 1 which means that people's choices are changed without them knowing that they are being influenced. Yet, if changes in behaviours are occurring at a non-conscious level then it is unlikely to bring about any substantive shift in people's value structure needed to solve society's major problems in the long term (Goodwin, 2012). Conversely, if nudges work, it might be the case that they could lead to sustained

behavioural change if they deliberately targeted conscious processing over a long period of time. Furthermore, if the theoretical basis on which nudge is founded is problematic, then this will raise implications for ethical issues around free choice and whether or not this is infringed upon through the use of nudges. Consequently, a sound conceptual framework is a pre-condition for assessing the effectiveness and acceptability of nudge as an intervention strategy (Baldwin, 2014).

In an attempt to address the two critical issues with the DST, the first objective of this thesis is to propose an alternative theoretical framework to account for the current nudge evidence-base. Based on the proposed framework, the idea is to apply a new conceptualisation to understand lifestyle behaviours such as poor diet, physical inactivity, alcohol overconsumption and tobacco use. These four main risk factors were selected for investigation due to their significant role in development of non-communicable diseases (NCDs). NCDs, principally heart diseases, stroke, cancer, diabetes, and chronic lung diseases are responsible for almost 70% of the global death (WHO, 2017a). Since we make choices in our day to day lives based on heuristics and biases that potentially drive poor lifestyle choices, therefore, gaining evidence on the effectiveness of these interventions can help determine how to best target NCDs in the future. Having established the critical concerns regarding the current DST, and presented the broad details of an alternative theoretical foundation, the next section introduces the ethical issues concerning the nudge programme, in particular, with respect to its working mechanism.

1.4.Nudge and Autonomy

By definition, nudges are designed to influence choice behaviour so to increase the likelihood that certain options will be selected. At the heart of nudge philosophy is libertarian paternalism which preserve people's rights to freely choose the preferred option (Sunstein, 2016d; Thaler & Sunstein, 2008). The justification for nudge is that the choice architecture is never neutral because choices have to be presented in one way or another (Thaler & Sunstein, 2008). This approach is designed to spare policy makers any ethical concerns, because a certain choice behaviour is being encouraged by the state that promotes a certain value-based lifestyle approach, yet preserves the right of an individual to do otherwise (for an in depth discussion see Osman, 2016). With regards to nudges, the concern is whether they easily allow people to make their own decisions; thus preserving their right to choose. Debates have arisen because some argue that the way in which nudges operate in changing choice behaviour is

predominately influenced by exploiting people's cognitive biases without their awareness (Ashcroft, 2013; Blumenthal-Barby & Burroughs, 2012; Bovens, 2009; Dworkin, 2012; Osman, 2016; Rebonato, 2013). Ultimately, if one is choosing an option without any awareness of how and why it is being chosen, it would be hard for people to opt out even if they wanted to. For this reason, a persistent criticism of nudge is infringement upon autonomy (Hausman & Welch, 2010). According to Dworkin (1988, p. 105), autonomy is "... *understood as a power of self-determination*". A second-order capacity of a person is to critically reflect upon one's first-order preferences, values, and the ability to identify or change in light of these preferences. Put simply, autonomy is an individual's capacity to self-govern, and to make their own choices free from coercion.

Proponents of nudge argue that human beings have limited bandwidth – that is the cognitive capacity to consciously process all the choices they make is limited; hence, nudges actually enhance autonomy because it gives people the freedom to focus on other matters that are more important in people's life (Sunstein, 2016a). In this light, Felsen and Reiner (2015) argue that nudge improves autonomy if it counteracts a bias as it promotes a choice that is in line with one's higher-order desires. Similarly, Sunstein (2013) argues that the strongest objection to autonomy is welfare, that is, infringement of autonomy may not be a serious problem as long as we are guided towards welfare-promoting ends. The underlying support for these arguments is based on the assumptions that the choice architecture should be primarily focused on facilitating individuals' pursuit of their own goals and there is an acceptably low opt out cost, so long as the intervention meets the conditions of transparency. In reality, however, these conditions are rarely satisfied, because firstly the choice architect often face the knowledge problem as they do not possess information about the chooser's "true" preference hence they will inevitably impose value substitution based on making some presumption about the chooser's higher-order desire. Secondly, the way nudges work involves bypassing the chooser's cognitive capacities for reasoning making her less in control of her actions (Bovens, 2009). This means it is difficult for the choosers to opt out of the arrangement if they wanted to. Therefore, for individuals to exercise autonomy over their own choices, they should be able to discern the implemented intervention, that is, the intervention should be transparent.

Following this, the second objective of the thesis is to empirically examine the extent to which nudge undermines autonomy: 1) If nudges do not align with the chooser's higher order desire

- the welfarist argument; and 2) If people cannot discern the nudge intervention and how it is used to change their behaviour - the transparency argument. The following subsections will explicate these two arguments in further detail.

1.4.1. The Welfarist Argument

Value Substitution. The objective of nudge is to steer people's behaviour in a welfare-promoting direction as judged by themselves but because the way some of the nudges work is by influencing people in a non-transparent way, for example, implementing default options, altering plate sizes, and visual illusions to reduce speed. These nudges deprived people of the ability to make their own choices according to their value as they do not engage in people's cognitive capacity for reasoning and deliberation. Instead, the choice architect selects among the individual's inconsistent preference based on some "general presumptions" to determine an individual's "true" preference" (Whitman & Rizzo, 2015). The problem with choice architects, including government officials, is that they lack sufficient information about the chooser's genuine wants. The fact that people are susceptible to nudges suggest that their preferences are less well defined (Sunstein and Thaler, 2003) and cannot be determined through simple observation of their choices. So, for this reason the choice architect will inevitably impose their value in substitution of the chooser's, otherwise known as "value substitution". This means people might be trapped with the options that do not align with their higher-order desire. The prospect of value substitution as a governing philosophy therefore violates people's autonomy and threatens welfare by projecting one's interest onto another (Rebonato, 2013; White, 2013). In line with this, Sunstein (2013) states that whatever the origins of the objection to paternalistic government, the force of those objections should depend on whether paternalism from government threatens to reduce people's welfare or intrudes on people's autonomy.

In order to evaluate the welfare consequence of nudging in public policy, it is necessary to categorise each nudge according to its objective, that is, pro-self or pro-social. Pro-self nudges seek to counterbalance irrational behaviour in order to maximise individual welfare (e.g., automatic enrolment onto pension plans), whereas pro-social nudges primarily seeks to maximize social welfare (e.g., automatic enrolment onto organ donor registers) (Hagman, Andersson, Västfjäll, & Tinghög, 2015). This distinction allows evaluation of respective consequences from their implementation. Consider one of the most well-known nudges -

default rules, which work by automatically enrolling people into a programme without having them make a deliberative choice. These form part of the choice architecture and because there is always a default setting within a particular choice context - they are unavoidable. In the context of default options, two of the most commonly implemented nudges are pro-self default in pension enrolment and pro-social default in organ donation registration. Here, individuals are nudged into making the “right” choices by the government under a default opt-out system. Pro-self pension enrolment defaults, automatically enrolls individuals onto a pension scheme upon entering into the workplace. Whereas the pro-social organ donation defaults automatically enrol citizens as organ donors. In both these cases, opting out requires administrative efforts on behalf of the individual because to do so requires filling out necessary paperwork. However, it has been argued that there is little value in having a nominal right to opt out because nudge is effective by exploiting cognitive biases of the choosers. This means that due to the power of default, the ability to opt out would likely be subject to the same cognitive biases (Rebonato, 2013). The problem with the force of inertia is that people will find it difficult to opt out of unsuitable defaults (Sunstein, 2015c). So, what would this entail for autonomy and actual welfare? The welfare consequence of being defaulted into these schemes are subtly different. In the former, being defaulted into an inappropriate contribution rate would impact the level of future retirement savings because default options do not take into account different risk preferences and spending behaviours in a population. In the latter, welfare concerns the livings as they have to make a decision in relation to the deceased’s body after death which is dependent on whether there is a reason to respect the deceased’s wishes. Under default opt-out, they might find it difficult to identify a deceased relative’s beliefs about organ donation when there is no evidence of explicit consent (Whyte, Selinger, Caplan, & Sadowski, 2012).

Although defaults in these contexts have been seeing increases in the number of people on the organ donation register or in retirement savings scheme, there is no empirical work examining people’s judgement around the use of defaults in these pro-self and pro-social contexts. From a welfarist perspective, nudge undermines autonomy if it does not align with people’s higher order desire. In order to explore the welfare-consequences of implementing defaults, part one of the second objective of this thesis sets out to empirically determine the extent to which third party perceive people’s true preference to donate or to register under a retirement savings scheme are sensitive to defaults. Chapter 3.1 will outline how defaults are applied in these

contexts respectively and how lack of autonomy may lead to misalignment with the chooser's higher order desire. The findings for pro-self pension enrolment will be presented in Chapter 4 and pro-social organ donation registration in Chapter 5.

1.4.2. The Transparency Argument

Another concern about nudging in public policy making is their covertness. This is particularly important because a government's attempt to tackle societal problems may not be judged ethically acceptable if an intervention is covert or if it intrudes on an individual's life (House of Lords, 2011). This transparency criterion has important implications on policy measures upon which accountability depends on. Thaler & Sunstein (2008) have stated that these are indispensable safeguards. From an autonomy perspective, the less transparent the nudge is, the greater is the threat to autonomy (Bovens, 2009). For people to exercise their autonomy, a certain degree of transparency is required. A lack of transparency blocks a person's capacity for reasoning, and hence offends both autonomy and dignity (Sunstein, 2016d). As mentioned, non-transparent nudges, such as defaults, constrain people's ability to opt-out of unsuitable options. Although Sunstein (2016c) argues that people will reject nudges that they see as unfit, the problem arises if people are not aware that a default has been implemented and so it is difficult for them to reject it, even if they want to. In order to avoid a nudge that is either ill intended or unfit, one needs to know that one is being nudged in the first place.

Furthermore, nudges that lack transparency are more susceptible to the power of the manipulators because they can subvert the chooser's decision-making. "Subverting" involves influencing people's choice "*to the extent that it does not engage or appeal to their capacity for reflection and deliberation*" (Sunstein, 2016a, p11). In line with this, Hausman and Welch (2010) argue that "shaping" people's choice that excludes rational persuasion is considered manipulative. They contend that there is a difference between attempting to persuade someone by means of facts and valid arguments and attempting to take advantage of people's cognitive bias to get them to make a choice. Thaler and Sunstein (2008) attempted to address the issue of manipulation through one of their guiding principles: *Transparency*. Their notion of transparency was derived from Rawl's publicity principle which "*in its simplest form bans government from selecting a policy that it would not be able or willing to defend publicly to its own citizens*" (Thaler & Sunstein, 2008, p244). For this reason, subliminal advertising is

argued to have violated the publicity principle because it is invisible and thus impossible to monitor; whether or not there is any actual evidence for its existence (Osman, Lin, & Ashcroft, 2017). However, Barton and Grüne-Yanoff (2015) argue that the publicity condition is hypothetical as it only states that policymakers can defend policies they are willing to defend publicly, but it does not require them to actually defend them publicly. The publicity principle, therefore, does not seem to guarantee against manipulative use of nudges or the prospect that they might infringe autonomy. In an ideal world, transparency should allow the chooser to identify the intentional action of the choice architect. Bovens (2009) argues that nudges should not only satisfy “type interference transparency” (how the nudge will interfere with our agency) but also “token interference transparency” (each specific intervention is made transparent to the nudgee). Transparency at this level means manipulation with ill intent is also less likely to be effective (Sunstein, 2016d).

Although there are concerns that making nudges transparent would render it less effective (Bovens, 2009), previous research has found that disclosing a nudge does not undermine its effectiveness (Loewenstein, Bryce, Hagmann, & Rajpal, 2014). In fact, evidence suggests that the public generally prefer transparent nudges over opaque ones (Felsen, Castelo, & Reiner, 2013). Indeed, in the House of Lords (2011) report, Luc Bovens concluded that ethical acceptability do not require governments to explain that an intervention has been implemented, but to render an intervention as acceptable those being nudged should have the ability to discern its implementation. In line with this view, Hansen and Jespersen (2013) suggested that distinguishing between transparent and non-transparent nudges will distinguish the manipulative use of nudges from other types. These points all highlight the importance of retaining transparency in policies. In this regard, people’s perception of transparency provides information for ethical considerations related to the introduction of a policy, as well as the expected effectiveness and acceptability of the intervention, and whether people would be willing to change their behaviour under these conditions. Part two of the second objective of the thesis is to empirically examine the extent to which nudge undermines autonomy if people cannot discern the intervention and how it is used to change their behaviour. Before going further, we would require a better understanding of transparency since the publicity principle has been objected on the grounds of being hypothetical. From the standpoint that nudge infringes upon autonomy, it makes sense to take the view of the nudgee rather than the nudger since their decisions are being influenced without awareness. Nonetheless, transparency is a

subjective concept, that is, an intervention may be deemed transparent by the choice architect but can be perceived otherwise by the nudgee. Chapter 3.2 includes a further discussion on transparency with the aim of establishing an account of transparency from the perspective of those being nudged. Chapter 6 will present the empirical findings on public attitudes toward transparent and non-transparent nudges.

1.5. Summary

With the rise of nudging in government administrations, stricter scrutiny should be implemented. Governments should take steps to ensure that behaviour change interventions are based on concrete evidence. However, there is currently insufficient evidence to draw firm conclusions on the effectiveness of nudges. Due to a lack of randomized controlled trials at a population level and the diversity of nudge-type intervention designs, one area that draws attention is the theoretical complexities concerning the underlying mechanism of nudge. Two critical issues associated with the DST were identified which prompted a need to establish further theoretical rigor before evaluating the effectiveness of nudge. The first objective of the thesis is to propose an alternative theoretical framework to account for the current evidence-base of nudge interventions in the health domain; with regards to four lifestyle behaviours: poor diet, physical inactivity, alcohol overconsumption and tobacco use.

The second objective of the thesis is to explore the extent to which nudge undermines autonomy: 1) If nudges do not align with the chooser's higher order desire - the welfarist argument; and 2) If people cannot discern the intervention and how it is used to change their behaviour - the transparency argument. From a welfarist perspective, proponents of nudge argue that intrusion on autonomy may not be a serious problem so long as nudges influence people in a welfare-promoting direction. I have illustrated that the welfarist argument loses its ground as nudge is prone to value substitution. This motivates two empirical studies to examine third party judgements of people's true preference when their choices are defaulted in the domain of pension enrolment (pro-self) and organ donation (pro-social).

On transparency grounds, another concern that draws ethical issues toward nudge is its covertness. Governments' attempt to solve societal problems might be judged as ethically unacceptable if an intervention lacks transparency. Similarly, an intervention that lacks transparency is also more likely to be considered manipulative. Although Thaler and Sunstein have attempted to address this issue by introducing the publicity principle, this condition is

often judged as insufficient due to its hypothetical nature. Since transparency is not in itself self-defining, this means government officials may perceive transparency differently from those being nudged. Hence, there is a need to establish a notion of transparency from the perspective of the nudgee rather than the nudger. This motivated the third empirical study to investigate public attitude towards transparent and non-transparent nudges as a way of providing information on ethical issues concerning transparency and autonomy. In sum, the thesis aims to critically examine nudge on theoretical, empirical and ethical grounds in order to guide policy making in designing ethically justified nudge-type interventions that are based on robust designs and which are more likely to sustain behavioural change at a population level.

CHAPTER 2 – Nudge: Theoretical and Empirical Re-consideration

Chapter 1 introduced the concept of nudge and explained the need to examine the critical issues associated with the underlying mechanism of nudge. This is to allow evaluation of the current nudge evidence-base within the health domain. The main focus in this chapter is to pick up the critical themes laid out in Chapter 1. The first section will outline the critical issues concerning the DST. Based on this, I will explore why a connectionist approach offers a better account of the relationship between System 1 and System 2. In doing so, I will propose a single system framework which classifies nudges into two types: Type 1 and Type 2, based on quantitative rather than qualitative distinction. The second section will then evaluate the evidence-base for Type 1 and Type 2 nudges in the health domain with regards to four lifestyle behaviours: poor diet, physical inactivity, alcohol overconsumption and tobacco use. To further establish empirical rigor, the third section will outline the empirical issues associated with nudges in the health domain. Since randomised controlled trials are hard to introduce in the field, it is important to consider internal and external reliability in future empirical research so as to draw firm conclusions about the effectiveness of nudges. Building on the proposed framework, review of the evidence-base, and the empirical issues the final section will re-consider the nudge programme by taking into account all these findings.

2.1.Critical issues with Dual-System theories

One area which needs inspection is the theoretical framework on which nudges are built on, given that this forms the rationale for how they are supposed to operate (Baldwin, 2014). This section examines the current proposed psychological mechanisms that underpin nudges, and the problems associated with the theoretical framework adopted. Based on this, an alternative theoretical account of nudges is proposed as a way to reconsider the evidence base of nudge interventions in the health domain. Chapter 1 has already introduced the theoretical framework of nudges and the critical issues associated with the distinction between implicit and explicit processing. The two core issues in the evaluation of System 1/System 2 distinctions are: 1) the nature of the interaction between the two systems, 2) the lack of precision around the details of key distinctions between the two systems (Osman, 2016). This evaluation is designed to better understand how to target behaviour via nudges, having a better idea of the actual underlying mechanism that guides decision-making should reveal where nudges are likely to be most

effective (Grüne-Yanoff, 2016; Grüne-Yanoff & Hertwig, 2015). The following section will examine these critical issues in further details.

Issue 1. The nature of the relationship between System 1 and 2. Dual system theorists do not all make the same claims regarding the relationship between systems (Evans & Stanovich, 2013). Some claim that the systems are interactive (Sloman, 1996), some claim that they operate in parallel (Epstein, 1973, 1994; Evans & Over, 1996), and others claim that they operate serially (Gilbert, 1989); System 1 is the default system, and only later does System 2 kick in to monitor the outputs of System 1 (Evans & Stanovich, 2013; Kahneman, 2003) - known as the default-interventionist approach. If System 1 and 2 are dissociated and do not interact, then it makes better sense to try to directly tap into System 1 processes to generate behavioural change; this is the default system that gives rise to many of the core decision-making and reasoning processes that drive sub-optimal lifestyle choices. As such, nudge theorists and practitioners need to identify common sub-optimal behaviours exclusively driven by System 1 processes, which thus far the program of nudges has yet to do.

However, if it is the case that the two systems do interact, then does it still make sense to introduce an intervention solely on System 1? If not, then nudges may not operate in the way that they are intended. In other words, if the grounds on which they are based are theoretically problematic, this may explain why nudges are not as effective as purported to be. To explore Issue 1 further we consider the bat and ball task (Kahneman & Frederick, 2002) that Thaler and Sunstein (2008) used to illustrate the relationship between System 1 and 2. The task involves presenting participants with a simple description and question: “A bat and ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?”. Typically, a majority of participants will intuitively answer 10 cents (Kahneman & Frederick, 2002), but the correct answer is in fact 5 cents. Just based on this task alone, DSTs vary significantly in their explanations for this common error. One interpretation is that System 1 is invoked automatically and is the sole driver of the error (Kahneman & Frederick, 2002, 2005), this is the favoured interpretation of Thaler and Sunstein (2008). An alternative explanation is that System 2 is also in operation but fails to detect the error generated by System 1 (Kahneman & Tversky, 1982), and a third interpretation is that because System 2 is slower than System 1, System 2 detects the error but cannot intervene quickly enough to prevent the error being made (Gilbert, 1989; Stanovich & West, 2000). The role that System 2 plays in this task, in turn has

implications for how researchers develop methods to effectively ameliorate the error. Similarly, when extended to nudges, the reasons behind a poor choice, in turn will impact what appropriate intervention will be required to reduce the chances of it continually being made.

Issue 2. Lack of precision regarding the critical distinctions between the two systems. There are essentially three different proposals regarding the core qualitative difference between System 1 and System 2. One view is that they can be differentiated based on their demands on working memory (De Neys, 2006; Evans & Stanovich, 2013; Oppenheimer, 2008). Working memory is a system for the temporary holding and manipulation of information during the performance of a range of cognitive tasks. A second claim is that they differ depending on the extent to which metacognitive processes are invoked (Thompson, 2009). This broadly refers to explicit knowledge or beliefs of factors that affect the outcome of a cognitive operation. The third is that they vary in the extent to which representations are accessible (Kahneman, 2003); the ease (or effort) with which particular mental contents come explicitly to mind. Essentially, the greater the dependency on working memory, or metacognitive processes, or difficulty in accessing representations suggests System 2 is in operation, and the opposite applies for System 1.

There are three main problems with the apparent qualitative distinctions between System 1 and 2. The first is, and as has been highlighted by DSTs, that in actual fact the three qualities essentially reduce to one single factor, namely dependency on working memory (Evans & Stanovich, 2013). This in and of itself is not necessarily a concern, since it suggests strong compatibility between the theoretical claims. However, what is a concern is that dependency on working memory is not all or nothing. Modelling and empirical demonstrations of the way in which high order cognition relies on working memory is based on quantitative differences, not qualitative ones (e.g., Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007).

Second is that the purported core qualitative distinction between System 1 and 2, be that working memory, metacognition, or accessibility of representations, is used to explain why one system is implicit/automatic, and the other is explicit. As with working memory, measures identifying automatic and explicit processes often rely on speed of response which is a continuous measure. This means that there are a variety of judgments, choices and inferences that are made, some of which are faster than others (Osman, 2004, 2007, 2013, 2014a; Osman

& Stavy, 2006). The relative nature by which implicit and explicit responses are identified, means that it makes better sense to claim that some responses are faster than others, rather than some are automatic/implicit and others are explicit. Thus far no DST theorist has presented the necessary and sufficient conditions by which to identify automatic/implicit responses independently of explicit responses (Osman, 2004). Moreover, not all behaviours associated with System 1 are fast, and not all behaviours associated with System 2 are slow, and to accommodate this some theorists have proposed a four system framework (Klaczynski, 2001; Sun, Slusarz, & Terry, 2005).

Third, the same initial observations that led to the formation of DSTs; and used to theorize about the key qualitative property that determines the difference between the two systems, are re-described to form predictions of the same observations which identify differences between the two systems. Put simply, it would be akin to detecting that sometimes people make erroneous choices quickly, and correct choices slowly, and then theorize that this is because of two underlying systems one of which is fast, the other of which is slow; each of which differ based on ease of access of information. From this, a prediction is formed which outlines that when the slow system is being used people will make correct choices, but not when the fast system is invoked. This is referred to as a particular type of circular argument (self-dependent justification), which has been commonly found in the area of DSTs of decision-making, judgment and reasoning (Hahn, 2011). Given the serious concerns discussed around Issue 1 and 2, the following section will propose an alternative.

2.2.A Single System Account of Nudge

Many dual system theorists (Evans & Over, 1996; Evans & Stanovich, 2013); as well as critics of them (Osman, 2004, 2013, 2014b; Gigerenzer & Regier, 1996) agree that a fully dissociationist view of System 1 and System 2 is not adequate for capturing the complexities in which decision-making processes operate. In light of this and other serious concerns with dual-system frameworks, several single-system frameworks have been proposed (e.g., Kruglanski, 2013; Osman, 2004, 2007, 2013; Osman & Stavy, 2006; Simon, Snow, & Read, 2004). Building on these unitary system frameworks, we extend their proposals by suggesting that they reduce to a parallel constraint satisfaction (PCS) model (Simon et al., 2004). This is essentially a connectionist approach in which the spread of activation among nodes in the network is fully sufficient for the processing of an outcome, e.g. a choice, and the basis on

which a decision is made (knowledge, evidence, beliefs) is coded in the network through the pattern of weights among the nodes (Read, Vanman, & Miller, 1997). PCS processing is guided by the goal of maximizing consistency, which means the need to reconfigure, re-evaluate, update knowledge, evidence and beliefs from multiple (potentially conflicting) sources that bring about an outcome (i.e. judgment, choice, inference) so that both (representations and outcome) are in alignment (i.e. coherent). Where DSTs identify distinct types of processes that can be classified as System 1 or System 2, in a PCS model, variations in processing of information is predicted and modelled according to the degree of restructuring that needs to occur for coherence (between knowledge and behaviour) to be achieved (Simon & Holyoak, 2002).

The PCS single system framework is extended to help classify nudges into two types that differ according to the degree to which processing efforts are needed to maintain psychological coherence (See Table 1 for examples). Type 1 nudges target decision-making contexts which generate responses that are not typically accompanied by critical inspection to prompt reconsideration of the choices made. For instance, familiar consumer-based contexts such as supermarkets involving highly practiced patterns of behaviour leading to repetitive choices being made. So, in response, Type 1 nudges involve simple interventions such as re-arranging presentations of consumer items in food aisles to highlight options that would have ordinarily been ignored. Type 1 nudges minimally disrupt the choice context to prompt some adjustment in the way information is processed at the point of decision, but not enough that the decision-maker detects any dissonance between the nudged choice and their general value-system. Type 2 nudges aim to promote a sustained re-evaluation of the evidence base on which people make their choices and the choices themselves, by disrupting the coherence between the two. For example, information prompt placed at the bottom of stairwells to promote exercise by presenting the benefits of regular exercise as well as the harmful effects of a persistent sedentary lifestyle. Repeated exposure to information of this kind is designed to create dissonance resulting from the costs of maintaining poor habits and the benefits of changing them. Given the cognitive system's need for coherence, a re-evaluation and restructuring of knowledge representations is needed to bring in alignment the evidence base (knowledge/beliefs) and choice behaviour. It is worth reiterating here that the two types of nudges differ according to the amount of re-evaluation of information on which peoples' choices are made and actions taken based on it, it is not predicated on a difference between

qualitative differences in systems of thought. Based on this conceptualisation, the evidence bases will be critically examined in order to determine which type of nudge is shown to be effective.

2.3. Empirical Evidence in Health Domain

The proposed framework will have implication for evaluating the effectiveness of nudge because the distinction between Type 1 and Type 2 nudges allows for the classification of interventions according to their ability to bring in alignment the evidence base and choice behaviour, and in turn determine their respective effectiveness. This section examines a range of nudges (Table 1) implemented in four health domains based on the proposed distinction: 1) poor diet, 2) physical inactivity, 3) alcohol overconsumption and 4) tobacco use. This is not a comprehensive review of the evidence in the literature, but instead is a focused evaluation of core findings to illustrate the pattern of evidence in this area (see Appendix A for studies that were included in the review).

Table 1.

A summary of the evidence on the effectiveness of Type 1 and Type 2 Nudges.

Health Domain	Type 1 Nudges		Type 2 Nudges	
Unhealthy Diet	Smaller plate sizes to reduce consumption	Limited effect	Calorie labelling	Ineffective
			Traffic light labelling	Effective
Physical Inactivity	Footprints	Ineffective	Motivational posters	Mixed evidence
Harmful Use of Alcohol	Adopt straight glassware	Insufficient evidence	Correct social norm misapprehension	Mixed evidence
Tobacco Use	Shorter cigarettes	Ineffective	Health warnings on branded packs	Mixed evidence
			Health Warnings on plain packs	Effective

Poor Diet

A common Type 1 nudge approach to address overconsumption includes changes to visual cues in a food environment which may consist of the availability of certain foods, the variety of food assortments, size of food packages and portions, or shape/size of plates, glasses and bowls. These cues are often used to imply a consumption norm that helps regulate how much we eat or drink in a food environment (Wansink, 2004). In addition, many people adhere to the

norm of ‘plate clean’, which means that in a food establishment if the plate is large, and the portions match the size, people consume to the size of the portion on the plate, and not to the point of being sated (Schwarz, 1998). For example, depending on the plate size, up to 45% more food is consumed in a Chinese buffet setting (Wansink & van Ittersum, 2013). A comparison of 58 studies (6603 participants) in a recent Cochrane review found that people consistently ate more food when offered larger-sized portions, packages or items of tableware, than when offered smaller-sized versions (Hollands et al., 2015). Introducing Type 1 nudges that reduce plate size in food establishments was shown to have a reasonable effect in reducing intake, but was dependent on participants being unaware of the manipulation (Holden, Zlatevska, & Dubelaar, 2016). However, it isn’t clear that the intervention successfully generalizes to contexts beyond the one in which the nudge was implemented. Moreover, many factors can significantly influence our eating habits. Eating is often a social activity, and we take our cues as to how much to consume from our dining partners, as well as other distractions that might affect our consumption (e.g., watching television, watching movies at the cinema). This suggests that there are multiple countervailing factors that limit the scope of Type 1 nudges (Wansink, 2004).

In this context, as an alternative to Type 1 nudges, Type 2 nudges focus on improving the presentation of information on which people make their food choices, such as providing calorie counts on food menus to draw attention to both healthy and unhealthy options (Downs, Loewenstein, & Wisdom, 2009). However, evidence suggests that this has had limited effects in increasing healthy food choices (Loewenstein et al., 2012). One reason for this is that calorie labels do not provide an obvious reference point as to which specific options are best. In and of themselves, they do not reliably motivate people to systematically monitor and translate calorie counts to shift their choices over a sustained period of time. Indeed, a recent systematic review of the impact of reading calorie labels at the point of purchase or consumption had little to no effect on positively changing people’s choice behaviour (Kiszko, Martinez, Abrams, & Elbel, 2014; Sinclair, Cooper, & Mansfield, 2014). The review found that regardless of the length of the intervention, the Type 2 nudge was generally ineffective [4 week period (Elbel, Gyamfi, & Kersh, 2011; K. L. Webb, Solomon, Sanders, Akiyama, & Crawford, 2011); 2 month period (Holmes, Serrano, Machin, Duetsch, & Davis, 2013), 13 month period (Finkelstein, Strombotne, Chan, & Krieger, 2011)]. To deal with these concerns, traffic light systems have been added to make calorie information more salient and intuitively simpler to

interpret (Sacks, Rayner, & Swinburn, 2009; Sonnenberg et al., 2013). By using simple visual cues (e.g., red = highly fat food, amber = moderately healthy foods, green = very healthy foods) the signals directly connect calorie counts to their impact on health (House of Lords, 2011), and provide a relevant reference point (Liu, Wisdom, Roberto, Liu, & Ubel, 2014). An independent field study conducted by Ipsos Mori showed that 35% of customers of a major UK supermarket actively look at traffic light labels when they shop, and 92% of those find these labels easy to understand. Also, over the 12-week period, sales of food items with mostly green traffic lights grew to 46.1% whereas those with mostly red traffic lights decreased by 24% (House of Lords, 2011). Thus, in the UK alone several organizations (The National Institute for Care and Health Excellence (NICE); The UK Food Standard Agency) have strongly encouraged food manufacturers and food establishments to use text and traffic lights on food labels/menus because they support greater alignment between knowledge (i.e., calorie information) and choice behaviour (i.e., be eat healthily).

Physical Activity

As with poor diet, another major global issue is the significant decreases in regular exercise (Hallal et al., 2012). A simple method of increasing physical activity through a Type 1 nudge involves a point-of-decision prompt which uses visual cues in relevant contexts to encourage people to take the more active of two options (e.g. a choice between taking the stairs rather than the escalator). For instance, painted footprints on stairwells have been used to guide people to take the stairs over elevators. The evidence base for this is not encouraging. Findings show that, perversely, the method increased the selection of the less physical option (Åvitsland, Solbraa, & Riiser, 2017). Similarly, Type 2 nudges also use a point-of-decision prompt, but instead of covert visual cues, educational information is presented that highlights the benefits of regular exercise. This often involves placing posters at start of stairwells or by elevators/escalators that inform people about the calories they would burn or the net positive effects on their health (i.e. increased heart rate) (Andersen, Franckowiak, Snyder, Bartlett, & Fontaine, 1998; Blamey, Mutrie, & Aitchison, 1995; Brownell, Stunkard, & Albaum, 1980; A. Lewis & Eves, 2012; Marshall, Bauman, Patch, Wilson, & Chen, 2002; Nomura, Yoshimoto, Akezaki, & Sato, 2009; O. J. Webb & Eves, 2007). A recent review of the evidence reported that across 11 studies the improvement in observed stair use was 50% from baseline (Soler et al., 2010), but a second review reported an overall positive effect ranging between 0.3-10.6% (Nocon, Müller-Riemenschneider, Nitzschke, & Willich, 2010). A speculation in the difference

between these two reviews is that there is variability in the length of the intervention. The length of the intervention is often between 4-12 weeks, but with some notable exceptions [24 weeks (Kerr, Eves, & Carroll, 2001a), 9 months (Lee et al., 2012)]. Another potential explanation for the mixed findings is that the locations in which the nudge was implemented varied, and so it is hard to compare like for like; for instance, comparing returning populations taking the stairs at train station versus those at a shopping mall. To further uncover the precise reasons for the variability of Type 2 nudges effectiveness in this domain, the types of informational prompts have been evaluated. A specific message such as ‘7 minutes of stair climbing protects your heart’ was shown to be more effective than a general message such as ‘Stay healthy, use the stairs’(Puig-Ribera & Eves, 2010). It is also worth noting that studies examining Type 1 and 2 nudges typically involve point-of-decision prompts placed at stairwell with only one to two flights of stairs. In order to experience any significant impact on cardiorespiratory fitness women need to climb at least 6 flights of stairs daily (Boreham, Wallace, & Nevill, 2000), and men need to climb 25 flights to result in any significant improvement in aerobic power (Fardy & Ilmarinen, 1975).

Alcohol Overconsumption

The evidence base for nudges designed to reduce alcohol overconsumption only accounts for 7.3% of all behavioural intervention studies in the health domain (Hollands, Shemilt, et al., 2013), despite the severity of the problem (Magnusson, 2009). Akin to the Type 1 nudges used to reduce food consumption via altering the size of food containers, a similar rationale has been adopted in the context of alcohol consumption. This typically involves offering alcohol in tall, narrow glasses as opposed to short, wide glasses in drinking establishments (bars and public houses) (Wansink & van Ittersum, 2005). This is motivated by work showing that the rate of alcohol consumption is related to the shape of glassware, which is slower in a straight glass compared to a curved glass (Attwood, Scott-Samuel, Stothart, Munafò, & Campanella, 2012). A recent systematic review of studies examining the use of this nudge reported that there was not enough evidence to estimate the effect on reducing consumption (Hollands et al., 2015). In line with other Type 2 nudges discussed so far, a preferred method is to provide explicit information to generate behavioural change. For example, providing a more accurate idea of safe quantities to consume through the use of social norm cues. As social creatures, people are sensitive to majority influences and this can be a strong persuasive influence on behaviour (Bullers, Cooper, & Russell, 2001; Ennett et al., 2006; Pearson & West, 2003). A wealth of

evidence from social psychology shows that people behave in accordance to what their peers do (J. V. Wood, 1989), often as part of a group mentality; there is a strong drive to belong and be accepted by a group. Given that consuming alcohol is typically a social activity, the claim is that using social norm cues (i.e. the typical amount of alcohol a particular social group consumes) is a more efficient way of helping people regulate their alcohol consumption, by evaluating it relative to the consumption of their peers (Nishida, Akaoka, & Nishizawa, 1975). For instance, heavy drinkers often judge their alcohol consumption to be equal to or even less than their peers – even though it is substantially greater (Perkins, Meilman, Leichliter, Cashin, & Presley, 1999; they feel as if they can reasonably justify their behaviour by rationalizing that it is no different to their peers). To correct misapprehensions of social norms in a student population, several studies using self-reported survey responses have shown that Type 2 social norm interventions (through educational campaigns) implemented over a period of one year (Laura Gomberg, Schneider, & DeJong, 2001) and five years (Haines & Spear, 1996) have successfully reduced alcohol consumption. However, a different review of 66 studies which analysed alcohol reduction at 4 months post intervention found that the effect sizes were small; and unlikely to have meaningful benefit in practice (Foxcraft, Moreira, Santimano, & Smith, 2015). It is worth noting that when surveyed, students doubted the credibility of such educational campaign messages (Thombs, Dotterer, Olds, Sharp, & Raub, 2004). In addition, some have suggested that the effectiveness of the nudge needs to take into account campus sizes in which norm misperception may be harder to correct if “everybody knows everyone else” and thus students are more confident in their estimates of other’s drinking levels (Borsari & Carey, 2003). Moreover, it is possible that the average or typical norm used to compare drinking levels in these Type 2 nudges does not represent the ideal normative reference point (M. A. Lewis & Neighbors, 2006). In other words, feedback that involve best friends’ drinking rather than typical student drinking level would be more specific and may have a stronger influence, assuming the peers are actually consuming alcohol within healthy limits (Baer, Stacy, & Larimer, 1991; Borsari & Carey, 2003; M. A. Lewis & Neighbors, 2006).

Tobacco Use

Another serious problem is tobacco consumption, which kills around 6 million people each year (WHO, 2016). Type 1 nudges promoting smoking cessation have focused on increasing the availability of shorter cigarettes, however, a systematic review found that when compared to standard sized cigarettes, there was no overall reduction in tobacco consumption (Hollands

et al., 2015). Alternatively, a more common route is to adopt educational campaigns typical of Type 2 nudges. A comprehensive review by Hammond et al. (2006) found that smokers' knowledge of toxic constituents in tobacco smoke was very low even amongst smokers in affluent and educated countries in the world. This led to changes in legislation in many western countries where health warnings on cigarette packages were adopted as the most common means of increasing smoker's awareness of the risks of smoking (Hammond et al., 2006). It is now mandatory that consumers of tobacco products have a 'fundamental right to health information, including accurate information about the harms of tobacco use' (WHO, 2015). While this is a common practice for many countries across the world, the style of presentation of health information differs between countries making it difficult to evaluate the effectiveness of these messages on reducing consumption.

More recently, Type 2 nudges such as plain cigarette packaging which standardize the shape, colour, method of opening the package, as well as the health warning labels have been used. The aim is to fulfil several objectives that include reducing the attractiveness of consuming tobacco (Hammond, Daniel, & White, 2013; Hammond & Parkinson, 2009; Moodie & Mackintosh, 2013; Moodie, Mackintosh, Hastings, & Ford, 2011), restricting use of the packaging as a form of advertising and promotion, while at the same time increasing the size of the health warnings (Maynard, Munafò, & Leonards, 2013; Moodie et al., 2012; Munafò, Roberts, Bauld, & Leonards, 2011). A large review of 37 studies concluded that plain packaging was rated as less attractive and contained poorer quality products than branded packaging (WHO, 2017b). Empirical work looking at the impact of health warnings on tobacco consumption is still in its infancy, but the findings are promising, and suggest that they indeed reduce acute craving, and are often associated with more negative perceptions of smoking (Brose, Chong, Aspinall, Michie, & McEwen, 2014). In the short term, plain packaging has been shown to encourage cessation for up to 2 weeks (Moodie & Mackintosh, 2013). In the medium and long term, there is evidence to suggest that plain packaging decreased tobacco consumption for 6 months (Dunlop, Dobbins, Young, Perez, & Currow, 2014), and 12 months (Wakefield et al., 2015) post-intervention. Meanwhile, Australia, being the first WHO member to implement standardize packaging, has also seen a statistically significant decline in smoking prevalence as a result of this Type 2 nudge (Australian Government - Department of Health, 2016).

2.4. Empirical Issues Concerning Nudge

Having examined the available evidence of both Type 1 and Type 2 nudges, this section will evaluate the methodological issues concerning the implementation of nudges designed to promote health behaviours. As with any intervention designed to improve behaviour, the most reliable way to confidently make casual inferences about a manipulation and its possible effect is to compare it against a control condition (randomized control trials, RCTs). However, RCTs are hard to implement in the field, and so along with other factors, limits the ability to draw firm conclusions on the effectiveness of nudges. Beyond this, there are two further key points that needs to be taken into account when addressing future empirical research on nudge.

Internal reliability of experiments examining nudges. Internal reliability refers to the extent to which a measure is consistent within itself – namely it generates the same behaviour each time it is used within the same context. The Type 1 nudges reviewed above suggest that overall the evidence-base is mixed, and replicability of positive nudge interventions is hard to establish. This raises questions about what the reasons are for when Type 1 nudges do work, and why the effects are hard to replicate. Loewenstein, Bryce, Hagmann, & Rajpal (2015) have speculated that the limited effectiveness of Type 1 nudges results from a lack of deep insight into how the nudge is designed to influence their behaviour. Ashcroft (2013) proposed that the effectiveness of both Type 1 and 2 nudges in general may depend on the various heterogeneous motivations/value systems people have with regards to changing their behaviour towards a healthier lifestyle. In addition, the fact that nudges are highly context dependent (Kosters & Van der Heijden, 2015), means that some Type 1 nudges are less likely to work in some contexts over others, and a clearer understanding of the context in which they are implemented is also needed. Thus, a critical step in devising research programs around nudging in health domains is to establish the internal reliability of nudges over time at an epidemiological level, but also at the individual level. However, the research practices so far have yet to adopt methodological techniques that tackle any of these issues in depth (i.e. assessing motivation needs, levels of awareness of nudges, characterizing the contexts in which they are implemented) in order to better establish internal reliability.

External reliability of experiments examining nudge. External reliability refers to the extent to which a measure varies from one use to another. With most nudge field experiments, the difficulty is in reproducing the same conditions in different contexts under which the original

intervention was assessed. For instance, consider nudges designed to increase physical activity. The informational prompts have been used in various environmental settings such as libraries (Russell, Dzewaltowski, & Ryan, 1999), underground stations (Blamey et al., 1995), office buildings (Coleman & Gonzalez, 2001). In meta-analytic review of these studies (Andersen et al., 1998; Blamey et al., 1995; Brownell et al., 1980; Kerr et al., 2001a; Kerr, Eves, & Carroll, 2001b; Marshall et al., 2002; Nomura et al., 2009), none suggested that there was a consistent pattern of evidence across the different contexts that were studied. The positive impact on behaviour as indicated in these studies increased stair use over elevators/escalators varied from around 2% to 12%, but not controlling for length of time in which the measure was implemented – i.e. 1 month or 3 months. Similar inconsistencies have also been noted for nudges that extend beyond the health domain. For instance, nudges used to increase civic behaviours, such as recycling, volunteering, voting, petitioning, donating, debating, have shown that the variation in how long, and where they are implemented may explain why overall effect size is as low as 9% (John et al., 2013). Thus, with findings such as this there needs to be more efforts in standardizing the ways in which nudges are examined in the wild in order to establish external reliability.

Thus, from an empirical perspective the picture appears to be somewhat bleak with respect to establishing good evidence for the effectiveness of nudges in the health domain. The main problem being that it is hard to draw any firm conclusions as to their effectiveness in the long-term (i.e. positive change over a year or more), which should be the ultimate goal of assessing their effectiveness. Often because the studies are conducted in the field, as with many field based studies, it is hard to run studies on a large sample with proper controls, further, it is rare to find field studies that also carry out follow-ups to examine the effects of the nudges in the long term. This does not undermine the program of nudge *per se*, but simply that the evidence to date does not allow researchers to draw strong conclusions about its general effectiveness in generating meaningful positive behavioural change. Moreover, as noted, the limitations in drawing firm conclusions is not restricted to their effectiveness and reliability over time, but also in establishing the generalizability of positive behavioural change beyond the context in which nudges were implemented. Furthermore, the small effect sizes reported in empirical studies means that translating their positive results at a population level may render them less effective than typical social policy methods used (i.e. mandates, bans, taxes). Given that there is a growing list of international governments wanting to apply nudge to public policy on

important issues such as health and wellbeing, there is clearly a need to establish further empirical rigor in order to better establish the effectiveness of these behaviour interventions (Osman, 2016).

2.5. General Theoretical Re-Considerations of the Nudge Evidence-base

As discussed earlier in this chapter, a concern for the nudge program is that the theoretical foundations in which it is built on is problematic. The first section of this chapter has proposed that the types of nudges can be put into two broad categories which differ depending on the extent to which they promote a re-evaluation of information that informs better decisions (i.e. maximizing long term gains), so as to bring the new information and choice behaviour into greater alignment (greater coherence). This is in contrast to the position of Thaler & Sunstein (2008) and Sunstein (2014b, 2016c); that nudges differ according to the underlying differences between System 1 and System 2 thought operations. Building on the proposed framework and the evidence reviewed, one reason why Type 1 nudges seem to be ineffective and short lived is because they do not engage the decision-maker on any substantial level to re-examine the basis on which their decisions are made so as to meaningfully shift their choice behaviour. This is consistent with Loewenstein et al.'s (2015) claim regarding the level of insight that people have as to the underlying basis on which Type 1 nudges are designed to influence their behaviour. Indeed, without prompting people to think and acknowledge that they might be consuming less as a result of smaller dinnerware/glassware, any behavioural change is not likely to become sticky (i.e. habitual), or reliably generalize to other contexts outside of where the nudge was implemented. It has long since been known that habits require sustained and explicit association between situational cues and learned behavioural responses (Hull, 1943), often through repetition of a behaviours in the same context (W. Wood, Quinn, & Kashy, 2002) for the behaviours to generalize beyond them.

The evidence review for Type 2 nudges typically involve interventions that utilizes provision of explicit information directly connected to the pursuit of a clearly identified goal, which in turn is associated with a specific choice behaviour (e.g. reduce unhealthy eating, alcohol consumption, tobacco consumption); this has been used in the form of providing calorie information, peer group's alcohol consumption, or health warnings on cigarette packages. More to the point, Type 2 nudges seem to be effective in reducing poor health behaviours such as alcohol consumption (Haines, Barker, & Rice, 2003; Haines & Spear, 1996) and cigarette

smoking (Hancock & Henry, 2003; Linkenbach, Perkins, & DeJong, 2003; Hancock, Abhold, Gascoigne, & Altekruze, 2002) for a period equal to or greater than 12 months. The evidence shows that through repeated intervention over long periods, some Type 2 nudges (particularly those correcting misapprehensions of social norms), can lead to sustainable behavioural change over longer periods of time (i.e. over a year post intervention). Thus, in order to establish reliable methods that promote critical re-examination of one's values, attitudes and motivations, it is advocated that Type 2 nudges should be more frequently used, and over sustained periods of time (i.e. at least 6-12 month educational campaigns). The rationale for this is that unlike Type 1 nudges, Type 2 nudges typically encourage a form of re-evaluation of behaviour through explicit means, and this helps to maintain greater coherence between the information on which new choice behaviours are made coherently. It is speculated for this reason that the evidence-based suggests that they are relatively more effective at sustaining behavioural change than Type 1 nudges (see Table 1).

2.6.Summary

In conclusion, the motivation for this chapter was to better understand how nudges work. With that in mind, once this is achieved, then the better armed we (social scientists, policy makers, practitioner) are in designing ways of intervening on behaviours to achieve the best outcome for individuals that need and want it. This chapter argued that any meaningful change in behaviour arises from developing a consistently coherent basis on which people understand the reasons for their decisions and how they enact them. If, through nudges, we want to encourage people to help themselves, particularly in targeting serious problems around NCDs, we need to make the goal of helping oneself making better lifestyle choices a coherent and sustained approach. For theoretical and empirical reasons, this is best achieved through Type 2 rather than Type 1 nudges.

CHAPER 3: Nudge and Autonomy

Chapter 1 introduced the critical themes on which nudges are examined on theoretical, empirical, and ethical grounds. In Chapter 2, I explored the first objective of the thesis which was to examine the critical issues associated with the current underlying mechanism of nudge and propose an alternative single system framework to account for the current evidence-base in the health domain. This chapter will focus on the second objective of the thesis which is to empirically explore the extent to which nudge undermines autonomy: 1) If nudges do not align with the chooser's higher order desire - the welfarist argument; and 2) If people cannot discern the intervention and how it is used to changed their behaviour – the transparent argument. The first section will explore the welfarist argument by introducing the concept of “true” preference and explain how nudges such as default rules may not promote people's welfare. From this, I introduce two types of defaults based on the objective of the interventions: pro-self and pro-social; and explore the welfare consequences of implementing these in the context of pro-self pension enrolment and pro-social organ donation. This motivates two empirical studies which compare third-party judgements of an individual's underlying wishes and nudged choices along with a Bayesian approach. At the end of each section, I will outline the rationale for the corresponding proposed hypotheses. The second section examines the extent to which nudge undermines autonomy when influencing people's behaviour without their awareness. I will explain the important implications in preserving one's autonomy and to ensure the right level of transparency in policies, there is a need to take into account public attitudes towards interventions as these provide information for ethical considerations related to introducing a policy. This motivated the third empirical study to examine transparency as an independent variable and how public opinion may shift based on the transparency of interventions. Subsequently, I will introduce four factors which could influence public opinion with regard to transparency of interventions: ease of identification, willingness to change, perceived effectiveness, and acceptability; and their corresponding hypothesis.

This following section will explore part one of the second objective of the thesis which is the extent to which nudge undermines autonomy when it does not align with the chooser's higher order desire - the welfarist argument. In Chapter 1, I have already illustrated how defaults can potentially intrude on autonomy and threaten welfare on the grounds that 1) they may not represent what people would have wanted and 2) they are sticky, making it difficult for people to opt out of unsuitable options. With pro-self and pro-social defaults, I have briefly explored

the consequence of being defaulted in the respective retirement savings context and organ donation context. The following section will explain the notion of “true” preference, and delve further into the consequence of being defaulted in various contexts.

3.1. The Welfarist Argument - “True” Preference in Default

Sunstein and Thaler (2003) have argued that in many domains, people lack clear, stable, or well-ordered preferences because people are strongly influenced by the details of the decision context such as default rules and framing effects, and if we take these influences into account, then people’s preferences are less clear. But this is perhaps not so accurate because people’s preferences are formed during the process of a choice. A person may not have a well-formed preference until he/she engages in the decision-making process and gradually learns about his or her preferences. In this view, *“the person has a true, underlying set of preferences to be uncovered”* (Whitman & Rizzo, 2015). In fact, Mill argues that the individual *“is the person most interested in his own well-being”* and the *“ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by anyone else”* (Mill, 1863, p21-2). Therefore, having control over one’s life would help determine his/her own welfare.

Given that one of the key underlying principles of nudge is to steer people’s behaviour in a welfare-promoting direction according to their own (true) preferences, this “true” preference has not been uncovered by choice architects. Instead, the choice architect designs the choice environment based on some general presumption of what people might want. For this reason, default rules often run into ethical objections. At a superficial level, defaults are suggestions for what people ought to be doing. In a sense, this is not manipulative if it simply conveys information, but the justification for default becomes complicated when considering the psychological mechanism that makes it effective. When implemented in organ donation, more than 99% of the population remained in the organ donor register (ODR) leading to an enormous difference in donation rate (Johnson & Goldstein, 2003). In the retirement savings context, defaults lead to 95% participation within a few months of employment compared to 60% participants without the default (Beshears, Choi, Laibson, & Madrian, 2008). The effect of opting-in vs opting-out generated a significant number of registrations. There are four reasons that account for their powerful effect. First, they appeal to people’s inertia such that people will stick to the status quo (Sunstein, 2015a). That is, moving away from this is perceived as a loss. The idea of loss aversion states that loss looms larger than gain (Tversky & Kahneman, 1991).

In other words, loss has a greater impact than the equivalent gain. Second, default can bridge the gap between a good intention (e.g., donate one's organs, save for the future) and the effort needed to implement that intention into practice (i.e., psychological barriers) (Shepherd, O'Carroll, & Ferguson, 2014). Third, defaults provide the decision-maker with important signals from policy-makers as to what ought to be the appropriate behaviour in situations of uncertainty (Davidai, Gilovich, & Ross, 2013; Johnson & Goldstein, 2003; McKenzie, Liersch, & Finkelstein, 2006). Whilst libertarian paternalists are tempted to assign the seemingly effectiveness of a default to people genuinely wanting to enrol in a pension scheme or to become an organ donor, it is possible that people are prevented from opting-out due to the very same cognitive bias that nudged people into the default option in the first place (Rebonato, 2013).

Given the way defaults are proposed to work, the choice architect is thought to exploit people's tendency to inertia and procrastination because individuals are enrolled onto a programme without making some kind of conscious choice. Sunstein (2016d) contends that if a default is not presented in a salient manner such that those being nudged are aware of the intervention in place, then it is important to consider active choosing to ensure the actual expression of agency. In fact, Johnson, Bellman, and Lohse (2002) found that people's own preferences are closer to choices made under no-default options and default opt-in options, suggesting that it is possible that active choosing is closer to some sense of "true" preference. The debate between active choosing and default choices will depend on which best represent people's true preference (hence promote welfare) in a particular context. But first, there is a need to distinguish between two types of nudges each of which have different objectives, namely, pro-self or pro-social. This distinction allows the evaluation of the welfare consequence of implementing defaults. As aforementioned in Chapter 1, pro-self nudges seek to counterbalance irrational behaviour in order to maximise individual welfare (e.g., automatic enrolment to pension plans) whereas pro-social nudges primarily seeks to maximize social welfare (e.g., automatic enrolment onto organ donor registers) (Hagman et al., 2015). Given the successful uptake of defaults, what has yet to be shown is whether there are differences between defaults systems (opt-in, opt-out) in terms of their signalling of choosers' "true" preferences (i.e., promote their welfare). In the following sections, I will first outline the background to two different implementations of defaults: pro-self and pro-social defaults. At the end of each section, I will discuss the extent to which these defaults signal one's "true" preference based on the background literature, and

provide a rationale for the specific hypotheses that were tested in the empirical work in this thesis. In particular, the empirical studies will explore the relationship between true preferences and nudged preferences. Thus, the remainder of this section is organized as follows. The first section lays out the background literature and the hypotheses tested in a pro-self default scheme, and the second section does the same for a pro-social default scheme.

3.1.1. Pro-self Default in Pension Enrolment

While we are in fortunate times with increasing life expectancy, there are several economic and social problems we face as a result (OECD, 2013; World Bank, 2016). Crucially, people are not saving enough for their retirement, and a major concern is that many retirees may outlive their retirement savings (Hanna, Kim, & Chen, 2016; Shu & Payne, 2015). Though there are some countries such as Canada (Jog & Lee, 2016) and the Netherlands (Knoef et al., 2016) that are bucking this trend. What might be the reason for the general failure to properly prepare for one's retirement? Studies conducted in the US, China, Netherlands, and Germany have demonstrated a strong link between poor financial literacy and lack of planning for retirement (Anthes, 2004; Bucher-Koenen & Lusardi, 2011; Lusardi & Mitchell, 2011a, 2011b, 2014; Song, 2015; Van Rooij, Lusardi, & Alessie, 2011). In addition, people hold negative attitudes towards future perceptions of a retired self (Mutran, Reitzes, & Fernandez, 1997; Noone, 2010), and lack the motivation and confidence in making complex investment decisions (Ameriks, Caplin, & Leahy, 2003; Lusardi & Mitchell, 2011a; Mandell & Klein, 2007). The general message from these findings is that social and economic issues regarding retirement protection require targeting psychological barriers. For this reason, the UK government was eager to incentivise people to save as well as invest in tax-efficient strategies (e.g., Individual Savings Account (ISA)). In addition, the Pension Act 2008 was introduced which placed the responsibility on employers to automatically enrol, through defaults, all eligible employees into a workplace pension; taking effect in the UK in 2012 (Department for Work & Pensions, 2008). The impact of this has been positive, with surveys reporting enrolment into pension systems increasing from 41% to 60% (Horack & Wood, 2005). Similarly, automatic default enrolment systems in the US have also reported dramatic increases in employer pension scheme enrolments up from 20-40% to 90% (Choi, Laibson, & Madrian, 2004). As part of automatic enrolment, there is a second layer of defaults in which the employee has their funds automatically invested into an investment fund, known as a Defined Contribution (DC)

workplace pension. DC pension plans are becoming increasingly common forms of retirement income provision in the UK, in Hong Kong (HK), and in the US (Byrne, Blake, Cairns, & Dowd, 2007). For instance, the HK Government rolled out the Mandatory Provident Fund (“MPF”) in which the employer automatically enrolls their employees into a pension, and each enrolled member’s savings are defaulted into pre-specified funds (Financial Services and the Treasury Bureau, 2015). The rationale here, as with automatic enrolment, is that the complexity of the investment products available on the market makes it difficult for people to make informed choices (Office of Fair Trading, 2013), so providing a defaulted investment fund takes away the effort of making a complex choice that might otherwise prevent them from adequately saving for their retirement.

Examples of automatic defaults, such as the 401(K) pension plan which appears in the Internal Revenue Code taxation system of the US, have been around since the 70’s. Once an individual is automatically defaulted (opt-out) into a retirement saving scheme typically saving rates and investment portfolios are selected on behalf of the employee. On this point, there is evidence that employees will reject the defaulted option if they mistrust the intentions of their employer (Shu, Bang, & Weber, 2016). On the other hand, employees tend to assume that the investment they are defaulted into is the most appropriate for them (especially if it is endorsed by governments and employers). In practice, given that there are many fund options, it is unlikely that nuanced decisions are made at the level of specific defaulted investment funds that account for each individual employee’s personal preferences (Byrne et al., 2007). Moreover, successful uptake of enrolment into retirement saving systems does not necessarily reflect what people have a preference for, only that they are more likely to remain in a defaulted system when compared to a no-default system (Jachimowicz, Duncan, Weber, & Johnson, 2016). Furthermore, an automatic default opt-out system is argued to preserve freedom of choice because an individual always has the option to actively leave the system (Benartzi & Thaler, 2013; Sunstein, 2013; Thaler & Sunstein, 2008). However, as aforementioned, default sticks as it is designed in such a way as to make it harder to opt-out (administratively there is an effort in taking oneself out of an automatic enrolment system). So, what do these factors entail for the actual preservation of free choice? First, people’s agency in their choice behaviour matters (Osman, 2014a), so much so that they often perceive defaults as ways of rescinding personal autonomy (Chapman, 2014). Second, people need to evaluate whether the defaulted option is indeed appropriate for them, but quite often, as discussed earlier, people lack the confidence

and knowledge to choose between financial products. This means if people choose not to choose, then the effect of inertia and the power of suggestions will lead people to stick with the default option even when it does not fit their situation. In fact, more recent evidence has revealed that those automatically enrolled onto a pension plan were 6% less likely to continue to make contributions over time compared to those who actively participated (Burk, Huang, & Luoto, 2015). Taken together, these factors present potential challenges to the general use of defaults in retirement savings contexts.

Thus far, there has been no empirical work examining people's judgments around the use of defaults in the context of retirement savings, and in particular with regards to people's sensitivity to defaults in countries where a mandated system is in place (i.e., Hong Kong), and where there is still some free choice exercised over investment schemes for retirement (i.e., UK). Therefore, when it comes to signalling an investor's true preference to register in retirement savings scheme, this study investigates whether people perceive automatic default opt-out systems differently from active choice systems (free choice, default opt-in). It also considers the influence on perceptions resulting from actual enrolment systems that different samples experience (i.e., free-choice in the UK, mandated system in HK). The main set-up of the study involves three enrolment systems: 1) Mandated choice system - the investor is legally required to make an active choice as to register in a retirement savings scheme; 2) Default opt-out system - the investor is automatically defaulted into a retirement savings scheme with a choice to opt out of the scheme; 3) Default opt-in system - the investor registers in a retirement savings scheme under an automatically defaulted opt-in system in which an active choice is made to opt in to the scheme. It is speculated that under an active choice system (i.e., default opt-in and mandated choice), decisions made overtly would provide a stronger signal of the decision-maker's preference and intention. Under an automatic default option such as opt-out systems, decisions are made passively as opposed to actively, hence these signals are perceived as more ambiguous. Chapter 4 will empirically test the hypothesis that:

H1 (pro-self welfare). Due to cultural differences between the samples, the UK and HK sample will differ in their probability estimate of the general assessment of people's preference to register in the enrolment systems.

H2 (pro-self welfare). People recognise the consequences of the different framings of the decision to be a registered investor in the enrolment system. They, therefore, see

registration as differentially representative of an investor's true preference to register, in the following order: *Default opt-in* > *Mandated choice* > *Default opt-out*.

3.1.2. Pro-social Default in Organ Donation Registration

The shortage of transplantable organs remains a pressing global issue. In 2017, the US had over 116,000 people on the waiting list to receive an organ and it was estimated that 20 people die each day while waiting on the list (Organdonor.gov, 2017). Similarly, 6,388 people in the UK were on the waiting list and approximately 450 patients died while waiting on this list (NHSBT, 2017). Defaults have been implemented to target this worldwide societal problem. Several studies have shown that default opt-out systems have substantially increased registered donations (Abadie & Gay, 2006; McKenzie et al., 2006; Shepherd et al., 2014; van Dalen & Henkens, 2014). For instance, Rithalia et al. (2009) reviewed five 'before and after' studies in which donation rates in Austria rose from 4.6 donors per million population (pmp) to 27.2 donors pmp over a 5-year period. In Belgium, kidney donations increased from 10.9 to 41.3 pmp during a 3-year period, and in Singapore, kidney donations increased from 4.7 to 31.3 per year over a 3-year period. To date, there has been much interest in the effects of organ donation legislations, such as the effect of implementing defaults in different organ donation legislations to generate donors (van Dalen & Henkens, 2014), families' attitudes and beliefs to decide whether to consent to organ donation (Exley, White, & Martin, 2002), and individual's willingness to donate their own or their relative's organs (Mossialos, Costa-Font, & Rudisill, 2008). There has not, however, been any dedicated examination of whether there are differences between default systems (opt-in vs opt-out) regarding the strength of the signal of the donor's true preferences to donate. Default appears to be effective in increasing donation rates so far but what remains outstanding then is whether default reflects people's true preference to donate. Sunstein's (2015a) defence is that if publicity and transparency are guaranteed then there are far less threat to autonomy. However, this is conditional on the basis that people are made aware of the nudge and that people also genuinely want to be nudged into the default (i.e. become an organ donor). If the basis of this is founded problematic, then it will have implication on family veto rates because whether families consent to the donation of the deceased will depend on what signals are being perceived from their registration on the organ donation system (e.g., active choice or default opt-out).

Defaults and Veto Decisions. In most organ donation legislative systems, default or otherwise, include a clause that allows the final decision to donate to be made by family members (Den Hartogh, 2012). This means that in effect family members can veto the decision made by their deceased relative. In fact, families vetoing decision is claimed to be one of the leading reasons for the gap between supply and demand of organs (Abadie & Gay, 2006; Barber, Falvey, Hamilton, Collett, & Rudge, 2006). In 2010, NHSBT reported that more than 500 families have vetoed organ donations despite being informed that their relative was on the NHS Organ Donor Register (ODR); this translated into an estimated 1,200 people missing out on potential life-saving transplants (NHSBT, 2016). Several studies have examined the factors that influence families' decisions regarding the overruling of their deceased relative's wishes (Exley et al., 2002; Mossialos et al., 2008; Rosenblum et al., 2012; van Dalen & Henkens, 2014). There is good evidence to suggest that if the deceased relative had made their decision known to their family in advance (an unambiguous signal of preference), then the family is more likely to honour their deceased relative's wishes (DeJong et al., 1998; Radecki & Jaccard, 1997; Siminoff & Lawrence, 2002). Consistent with this, evidence from survey work showed that 96.7% of US citizens (U.S. Department of Health and Human Services, 2013) and 88% of UK citizens (NHSBT, 2015) would consent to the donation of a deceased relative's organs if the deceased's wishes were made known in advance to the family. In reality, however, the wishes of the donors are often not known to the family, as revealed by studies showing that less than half of Europeans and North Americans had raised the subject of organ donation with their family (Eurobarometer, 2007; Spital, 1995). From this, one speculation is that, just as the default option in a country's organ donation legislative system may signal to the individual what ought to be the appropriate behaviour (Johnson & Goldstein, 2003), it can also act as a signal to the family as to their deceased relative's true preference. In other words, when faced with decisions to consent on behalf of their deceased relative, defaults may also influence families' perceptions of the deceased's true preference to donate given that the deceased was registered to donate. This is because strength of preference and choice satisfaction vary between default systems (i.e. personally made choice in opt-in vs. externally made choice in opt-out); personally choosing leads to greater satisfaction in the choice made (Botti & McGill, 2006; Payne, Bettman, & Johnson, 1993). Therefore, strength of preference and choice satisfaction is likely to be perceived by families as weaker under a default opt-out system because it involves a passive choice to donate compared to a default opt-in system where an active choice to donate is made.

Why might this be the case? In an explicit consent system (i.e. default opt-in), consent is expressed through some overt communication which can be seen as providing reliable evidence that a decision was made. In addition, active choices of this kind are made through a free self-selection process thereby affirming one's personal agency in the choices one makes (Osman, 2014a). The meaning attached to the act of donation is seen as an altruistic inclination hence representing a stronger strength of signal of true preference to donate. This is different from a presumed consent system (i.e. default opt-out), where the absence of an objection is recorded as consent which can potentially mute the strength of a signal of one's true preference to donate. A default opt-out system also sends a signal to the potential donor and their family that organ donation is a socially preferred choice (or recommendation by policy makers) (I. G. Cohen, Lynch, & Robertson, 2016) rather than the individual's true preference. Indeed, Davidai et al. (2013) has shown that participants assign a lower value to the act of being a donor in the default opt-out system compared to the default opt-in system. Given that donors rarely communicate their donation wishes with their families, their true preference to donate are often inferred from the ODR which is likely to be weaker and more ambiguous when a passive choice is made. The consequence of this is that it adds uncertainty to the families' decision when deciding whether to donate their deceased relative's organs. As mentioned, in the event of uncertainty, families are more likely to refuse consent. Therefore, by implication, the stronger the signal of true preference to donate is, given the type of organ donation legislative system, the less likely it is that families will refuse their relatives' wishes to donate.

The literature review shows that inferences of true preference from choice behaviour play an important role in the donation process and, without greater understanding, we cannot accurately predict or evaluate the consequences of new policies that involve defaults (Bowels, 1998). In other words, the effectiveness of such policies would depend both on the preferences they induce or evoke (Sunstein, 1993) and, in turn, the inferences relatives make in these instances. Chapter 5 will investigate people's sensitivity to the influence of organ donation legislative systems under which registration was made across four systems: default opt-in (everyone is automatically a non-donor unless one registers to be a donor), default opt-out (everyone is automatically a donor unless one objects), mandated choice (everyone is required by law to

state in advance whether they are willing to be a donor), and mandatory donor (everyone is required by law to be a donor and there is no option to change this)¹.

Chapter 5 will empirically test the hypotheses that:

H1 (pro-social welfare). People recognize the consequences of the different framings of the decision to be a registered organ donor. They, therefore, see registration as differentially representative of an individual's true preference to donate, in the following order: Default opt-in > Mandated choice > Default opt-out > Mandatory Donor.

H2 (pro-social welfare). If Hypothesis 1 is supported, then families making decisions on the basis of their beliefs about the deceased's true preferences to donate will be perceived by participants as more likely to agree to donation in the order predicted by Hypothesis 1.

3.1.3. Bayesian Approach

In the empirical work that follows, a Bayesian approach has been adopted to infer people's perceived true preference in both pro-self default in the retirement savings context and pro-social default in the organ donation context. In Bayesian inference, probability is a way to represent an individual's degree of belief in a statement and allows one to determine the probability of the model given data. Bayesian, therefore, provides people the tools to update their beliefs in the evidence of new data, or in other words, a normative standard for belief revision. This is useful because we can combine new data with prior knowledges to make inferences that better reflect the real-world decision making. The Bayes' Theorem states that one's posterior belief in a hypothesis, h (i.e. one's true preference to register in retirement savings scheme or donate in an organ donation), in light of the evidence, e (their registration on the system) is $P(h|e)$ which is a function of one's prior belief, $P(h)$, and the likelihood of observing that evidence if one's initial hypothesis was true $P(e|h)$, as opposed to if it was false, $P(e|\neg h)$.

Using this approach, it is possible to explore the consequences of people being defaulted into a retirement savings scheme or organ donation system. To infer an individual's true preference,

¹ The mandatory donor condition was added to the pro-social default study but not the pro-self default study because of differences in the contexts. In mandatory donor system, this is akin to being forced to donate and if people are giving sensible judgements then this option should be preferred the least.

third-party judgements² are used to derive the diagnosticity of an item of evidence. From a Bayesian perspective, the diagnosticity of an item of evidence, here registration on the enrolment system/ODR, is determined by the likelihood ratio –its informativeness regarding the hypothesis in question which is typically viewed as a measure of quality of the evidence itself. The likelihood ratio is the ratio of obtaining that evidence in the event that the underlying hypothesis is true (the - ‘hit’ rate) as opposed to obtaining that evidence in the event that the underlying hypothesis is false (‘false positive’ rate), $\frac{P(e|h)}{P(e|\neg h)}$. Where the likelihood ratio is 1, the evidence is just as likely to occur whether the hypothesis is true or false and is therefore maximally uninformative. A greater likelihood ratio, $P(e|h) > P(e|\neg h)$ will be associated with an increase in belief in h . The intuitiveness of this result is in line with Laplace's (1814/1951) observation that the probability calculus is “nothing but formalized common sense.” In line with this, there is evidence to suggest that people are sensitive to subtle sources of influences on the diagnosticity of evidence (Harris, Corner, & Hahn, 2013; Lagnado, Fenton, & Neil, 2012).

3.1.4. Summary – Pro-social and Pro-self nudges

So far, this chapter has illustrated how defaults can potentially intrude on autonomy from a welfare perspective: 1) they may not represent what people would have wanted and 2) they stick, making it difficult for people to opt out of unsuitable options. With pro-self and pro-social defaults, I have explored the consequence of being defaulted in the respective retirement savings context and organ donation context which motivates the empirical studies to examine third party judgement with regard to people’s true preference under different systems of registration (i.e., default opt-in, default opt-out, mandated choice, etc). This will be investigated along with a Bayesian approach, in which the empirical work and the analyses are presented in Chapter 4 (pro-self default) and Chapter 5 (pro-social default).

² In the pro-social default context, third party judgements reflect a real-world situation when one would make a decision on behalf of the deceased. To maintain this consistency, third party judgements were also adopted in the pro-self default context. This difference should not in principle impact the findings from the Bayesian analyses that are conducted; though there may well be subtle differences in the interpretation of the results. This is discussed in the general discussion section.

3.2. The Transparency Argument: Public Attitudes Towards Nudge

In this section, I will explore part two of the second objective of the thesis which is the extent nudge undermines autonomy when people cannot discern the intervention and how it is used to change their behaviour - the transparency argument. Nudges are less explicit compared to typical social policy tools such as mandates, taxes and bans which more explicitly and heavily steer people to behave in a manner that maximizes their own and society's good (i.e. seatbelt laws, fat tax, smoking bans). However, if nudge is included as part of a social policy tool, then disclosing the intention and implementation is a prerequisite, at least for any democratic society upon which accountability and responsibility are important. But because the mechanism by which nudges work is by triggering behaviour change without the chooser's awareness, this type of nonconscious influence poses profound questions regarding the manipulative nature of nudge (Bovens, 2009; Vallgård, 2012). Manipulation is objectionable because it involves some underhanded interferences with the ways in which people see their options and consequently make their choices. In this respect, manipulation violates autonomy even if the goal is to make people better off (Wilkinson, 2013). In fact, scholars have found that citizens object to manipulations, such as visual illusions to reduce speeding (Jung & Mellers, 2016) and subliminal advertising (Reisch & Sunstein, 2016). This means that the ethical acceptability of nudge to a certain extent depends on whether a nudge interferes with one's life and whether it is covert (Tomer, 2017).

Previous findings have shown that people judged the acceptability of a nudge and its moral implications in different ways. For example, Felsen et al. (2013) found that individuals tend to prefer overt rather than covert influence, and those who favoured overt nudges also perceived the nudge as more "authentic". Similarly, Jung and Mellers (2016) and Sunstein (2016c) found that System 2 nudges that target deliberative processing were perceived as more popular than System 1 nudges that target automatic processing. These findings suggest that public attitudes can be influenced by the transparency of an intervention but it is worth noting that none of these studies have directly assessed transparency as a measurement but rather the transparency of these interventions were defined by the type of nudge presented (e.g., System 1 or System 2). If as shown, people are less favourable of non-transparent/covert nudges then it makes better sense to examine the extent to which public attitudes are influenced by the transparency of different nudges. As mentioned, the concept of transparency in itself is not self-defining and in the context of nudging, those being nudged can perceive transparency differently from those

that designed the nudge. To successfully select and implement nudge, policy makers need a psychological understanding of how transparency is being perceived by those being nudged and base the implementation on which factors shape their attitude towards nudge. These factors include ease of identifying the intervention, the expected effectiveness and acceptability of the intervention, and whether people are willing to change their behaviour. All of these helps to determine public attitudes towards different types of nudge interventions (i.e., Type 1 and Type 2 – as I have specified them) which is a central issue for ethical considerations related to the introduction of a policy.

To provide empirical support for the transparency argument, the thesis empirically investigates the extent to which public attitudes towards these four factors are dependent on the transparency of an intervention, as well as the rationale presented for nudge-type interventions and the agent that proposed the intervention. These predictors are discussed in more detail below. Given that, in the introductory chapters of this thesis, I have evaluated the effectiveness of nudges in the health domain, to ensure consistency, the empirical study examining public attitudes also focuses on four health contexts, namely, smoking, food, alcohol, and exercise. As a relevant addition, the fifth context in which public attitudes was chemes. In the following section, I will outline the background to the four factors and the corresponding rationales for the hypotheses that are empirically tested in Chapter 6. But before going further, I will redefine the notion of transparency from those being nudged as this form the basis on which nudges infringe upon autonomy.

Transparency Redefined. To combat manipulation, Thaler and Sunstein (2008) proposed Rawl's Publicity Principle which in essence bans government from selecting a policy that it would not be able or willing to defend publicly to its citizens. However, it has been argued in Chapter 1 that this is an insufficient condition in guaranteeing against non-legitimate government manipulation because it is a hypothetical condition that does not require policy makers to actually publicly defend their policy. As with any social policy tools, the public should be able to review and scrutinize nudges no less than government actions of any kind. For this reason, Thaler and Sunstein's definition of transparency needs to be redefined from the perspective of those being nudged. To ensure the right level of transparency involves telling people about a nudge intervention directly or for a perceptive person to discern for themselves that a nudge intervention has been implemented (House of Lords, 2011). In line with this,

Bovens (2009) suggested the principle of token-interference transparency which states that a watchful individual should be able to discern the intention of the change in the choice architecture and to blow the whistle if he or she sees the nudge as unfit. Building on these, Hansen and Jespersen (2013) provided a more comprehensive definition of transparency, which states that a transparent nudge should be provided in such a way that the intention behind it, as well as the means by which behavioural change is pursued could reasonably expect to be transparent to the agent being nudged, as a result of the intervention. Conversely, a non-transparent nudge works in a way that the citizen in the situation cannot reconstruct either the intention or the means by which behavioural change is pursued.

Based on this, the key components of transparency of an intervention can be extracted to include discern the “*implementation*” and “*intentions/means*” of behaviour change. The former refers to identifying the actual intervention used to change people’s behaviour, whereas as the latter refers to identifying the psychological method used to change people’s behaviour. Taken together, it is possible to offer a definition from the perspective of those being nudged on two ends of the transparency scale from definitely transparent to definitely non-transparent so as to empirically test the extent to which individuals can discern that an intervention has been implemented:

A transparent nudge works in a such a way that anyone can easily identify the actual psychological method used to change their behaviour, as well as easily identify how their behaviour is changed by it.

A non-transparent nudge works in a way that no one can identify the actual psychological method used to change their behaviour, and no one can identify how their behaviour is changed by it.

This distinction is important because transparency may serve as a basis for evaluating nudges as working by manipulation (Hansen & Jespersen, 2013). Certainly, if an agent can identify the nudge, then she possesses the autonomous power to either accept or reject the nudge. In turn, this will mitigate the likelihood of welfare-reducing nudges that threaten personal autonomy. Given the above definition, it can be hypothesised that:

H1 (transparency). Across the five domains, the public will find it easier to identify the actual behavioural intervention and how their behaviour will be changed by the psychological method in the transparent condition and in turn, more accurately classify these as transparent interventions.

Willingness to Change Behaviour. What is currently missing from the nudge programme is people's internal willingness to change. Philosopher Ervin Laszlo (1989, p25) suggested that “*only by redesigning our thinking and acting, not the world around us*” can we solve the causes of our problems. As aforementioned, if a nudge intervention is going to work, then it is most likely work by targeting the conscious system. Type 1 nudge, for example, typically focuses on external influences which have limited effects. Eating less because the plate is smaller does not necessarily transfer to eating less at home. What matters is people's willingness to change their behaviour as this reveals their willingness to engage in the particular context in which the nudge is present. This reveals the intrinsic value they assign to the act of changing behaviour and whether it aligns with their higher purposes (e.g. wanting to save more, eat healthily). Indeed, there is evidence suggesting that the ability to deliberately engage in a behaviour will encourage a shift towards intrinsic value (Hogg, 2011). Deliberation requires consciousness which works on the basis that the intervention is transparent so that people can detect how and why an intervention is used to change their behaviour as well as to determine whether the behaviour change aligned with their value. It can thus be hypothesised that:

H2 (transparency). Across the five domains, the public will be more willing to change their behaviour when presented with transparent nudges compared to non-transparent nudges.

Perceived Effectiveness. Bovens (2009) suggested that nudges typically work best when people are unaware of it. For example, if we tell students that the order of food has been rearranged for dietary purposes or if we explain to employees that they are being defaulted into the pension scheme because of their inclination to inertia, then the intervention is likely to be less successful. Similarly, Sunstein (2016d) raised the question that if people can discern the nudge, and are explained why and how the nudges are implemented, would the nudge still be deemed less effective? There is evidence suggesting that revealing the default option does not undermine its effectiveness (Arad & Rubinstein, 2015; Loewenstein et al., 2014). These findings suggest that people generally prefer nudges that are transparent and effective, and so

there are empirical grounds to believe that people will generally judge a transparent intervention as more effective than a non-transparent one.

H3 (transparency). Across the five domains, the public will generally perceive an intervention as more effective in the transparent condition than the non-transparent condition.

Public Acceptability. To the extent that the acceptability of nudge have been extensively debated (Hansen & Jespersen, 2013; Hausman & Welch, 2010; Marteau et al., 2011), there has been a lack of primary research in the public acceptance of nudge interventions (Marteau et al. 2011). The level of public acceptance is relevant in determining how ethical a nudge is perceived to be. There is some evidence suggesting that transparency of an intervention – i.e. whether it is overt or covert – also has an influence on public acceptability with interventions that operate via conscious processes (i.e., overt nudges) gain higher acceptability (Felsen et al., 2013). However, Petrescu, Hollands, Couturier, Ng, and Marteau (2016) were not able to replicate these findings. Instead they found no evidence to suggest that highlighting the conscious or non-conscious processes of interventions affected acceptability. They argued that the difference in the findings are due to the use of different wordings to describe non-conscious process, for instance, they used the wording “people will not be conscious (i.e. not aware)” whereas Felsen et al. used the wording “subconsciously driven bias”. Further, Pestescu et al.’s study assessed acceptability directly, whereas Felsen et al. used a proxy measure which asked participants whether or not they prefer working for a company that employs the intervention. Due to the lack of consensus in these findings, further empirical work is needed to establish whether the transparency of an intervention influences the public’s judged acceptability of the nudge. Based on this, it can be hypothesized that:

H4 (transparency). Across the five domains, the public will generally perceive an intervention as more acceptable in the transparent condition than the non-transparent condition.

3.2.1. Other Predictors

Rationale of Interventions. Is it manipulative if we do not tell people how an intervention influences their behaviour? The obligation to inform people about the specific mechanism that makes nudges effective is also one of Bovens’ (2009) concerns. Even though, previous studies have presented participants with simple rationale for how the specified intervention works (Felsen et al., 2013; Petrescu et al., 2016), little is known about whether revealing the

psychological mechanism will influence public opinion with regards to the ease of identification, perceived effectiveness and acceptability of an intervention, and to the extent it will affect people's willingness to change. To test for this, the psychological mechanism of nudges will be presented positively (how it does work), negatively (how it does not work) and both positive and negative. Intuitively, people would be more persuaded by psychological mechanism of nudge that has a positive rationale since people are more supportive of nudges when there is evidence to suggest that it is effective (Sunstein, 2016c). It can be hypothesised that:

H5 (transparency). Public opinion with regard to ease of identification, perceived effectiveness and acceptability of an intervention, and their willingness to change will be influenced by the presentation of rationale for these interventions and judgements will follow the direction of: positive > positive + negative > negative.

Behaviour Agent. The question of who is nudging also influences public opinion. The intention and motivation of nudger/choice architect can be perceived differently, for example, people tend to perceive that default options are implicit recommendation by policymakers (McKenzie et al., 2006). Typically, support for nudges diminishes when people distrust the motivations of the choice architects or they fear that they would end up with outcomes that are inconsistent with their value (Sunstein, 2016b). An international survey conducted by Ipsos MORI on 18,500 adults across 24 countries found while the majority are in support of many specific interventions, as many as half of them say that they do not think the government should get involved in people's choices. A third of the public would endorse tougher interventions but think that the state should not get involved in people's choices around how they eat, save, or live sustainably (Branson, Duffy, Perry, & Wellings, 2012). It is the intentional component of the nudge that is of greater importance in explaining why manipulation is viewed as wrong (Wilkinson, 2013). Given that the intention of the agent can be perceived differently depending on who proposes the nudge, the study will explore this by introducing three types of agents, namely, government officials, advertisers and researchers. It is speculated that, researchers in this case act as the "neutral" group because they are seen as potentially having less of an obvious agenda to introduce an intervention as compared to advertisers and policy makers. Therefore, it can be hypothesized that:

H6 (transparency). Public opinion with regard to ease of identification, perceived effectiveness and acceptability of an intervention, and their willingness to change will be influenced by

the behaviour agent that proposed the intervention, such that judgments overall will be most favourable towards scientists proposing nudges as compared to advertisers and policy makers.

3.2.2. Summary – Public Attitude Towards Nudges

This chapter explored the importance of transparency in preserving a person's autonomy to act in accordance with his/her preference. Even though previous survey findings have revealed that people generally prefer transparent over non-transparent nudges, none have explicitly determined people's perception of transparency and how that impacts other judgments such as ease of identifying the transparency of interventions, their willingness to change their behaviour, as well as how acceptable and effective the nudges are perceived to be.

As argued in this chapter, transparency can be perceived differently by the nudger and the nudgee. If transparency is needed to preserve people's autonomy then those being nudged should be able to discern the intention and means of behaviour change for themselves. If they cannot, then the nudge, at least for them, is not transparent. Having redefined transparency, the study will explore public attitude related to four factors: ease of identification, willingness to change, perceived effectiveness and acceptability; and the extent to which these factors are dependent on the transparency of interventions, the agent that proposes the intervention and the presentation of rationale behind these interventions in five contexts (i.e., smoking, food, alcohol, exercise, and banking). To summarize the hypotheses, across the five contexts it is predicted that people will find it easier to identify the intervention and will more accurately classify it as transparent, perceive it as more effective and acceptable and are more willing to change their behaviour in the transparent condition as opposed to the non-transparent condition. Their attitude towards nudge will also be affected by the agents that proposed the intervention and the rationales presented for the intervention. The implication of these findings will provide relevant information for thinking about the support for nudge type interventions and its appropriateness as a social policy tool.

CHAPTER 4: Underlying Wishes and Nudged Choices in Pro-self Context

Chapter 3.1 examined part one of the second objective of the thesis – the welfarist argument – under which nudge undermines autonomy in pro-self and pro-social contexts. I have argued that choice architects have insufficient information about the chooser’s underlying wishes and because of this, they inevitably impose value substitution. In this Chapter, I will provide the empirical evidence for the welfare consequence of implementing pro-self defaults in pension enrolment. The present study (Experiment 1a, 1b and 2a, 2b) investigates the differences in judgement between choices made under different nudged choice contexts. The main focus of this chapter will be on providing empirical findings on people’s judgement of an individual’s “true” preference to register in a retirement savings scheme, based also on the application of a Bayesian approach. This will be investigated with respect to two samples: UK and HK.

First, given that different retirement savings schemes are implemented in these countries, the study first explores the differences in perceptions between the UK and HK samples by eliciting their prior belief, $P(h)$. This is the probability estimates of the general assessment of people’s preference to register before taking into account evidence ‘ e ’ (i.e., their registration in the enrolment system). Given their current experiences of different enrolment systems in their country, $P(h)$ is explored through two measures: 1) by comparing judgements of $P(h)$ between UK and HK samples in Experiment 1a and 1b; and 2) investigating the effect of the presentation of different enrolment systems on the elicitation of $P(h)$ in Experiment 2a and 2b. The prior belief in Experiment 2 was elicited after presentation of the respective retirement savings schemes to see if these influence people’s judgement of priors. This set-up is different from the pro-social context because Experiment 2 was initially conducted as a pilot study. But due to the fact that there was a difference in the elicitation of priors as a result of the order of presenting the schemes, the results from Experiment 2 was also included to offer additional insights.

Second, in order to infer an investor’s true preference, third-party judgements are used to derive the diagnosticity of an item of evidence which is determined by the likelihood ratio. This is the ratio of obtaining that evidence in the event that the underlying hypothesis, h , is true (the investor’s true preference was to register to invest in a retirement savings scheme, ‘hit’ rate) as opposed to obtaining that evidence in the event that the underlying hypothesis is false (the

investor's true preference was *not* to register to invest in a retirement savings scheme, 'false positive' rate), $\frac{P(\text{registration}|\text{want to invest})}{P(\text{registration}|\text{DON'T want to invest})}$. Based on this, the perceived diagnosticity of an investor's true preference to register in a retirement savings scheme is investigated across three experimental conditions, namely default opt-in (everyone is automatically a non-investor unless one actively registers in the enrolment system), default opt-out (everyone is automatically an investor unless one objects to register in the enrolment system), mandated choice (everyone is required by law to state in advance whether or not they are willing to register in the enrolment system).

The specific hypothesis investigated here is:

H1 (pro-self welfare). Due to cultural differences between the samples, the UK and HK sample will differ in their probability estimate of the general assessment of people's preference to register in the enrolment systems.

H2 (pro-self welfare). People recognise the consequences of the different framings of the decision to be a registered investor in the enrolment system. They, therefore, see registration as differentially representative of an investor's true preference to register, in the following order: *Default opt-in > Mandated choice > Default opt-out*.

Given the set-up of the study explained above, the present study will be presented in the format of Experiment 1a & 1b; and Experiment 2a & 2b, where "a" represents the UK sample, and "b" represents the HK sample. For all four experiments ethics approval was granted by Hang Seng Management College Research Committee.

In all four experiments, three enrolment systems (default opt-in, default opt-out, mandated choice) were set in the context of retirement savings schemes. In Experiment "a", the sample consisted of UK residents, in which a common investment option is the ISA. Though not exclusively a retirement savings scheme, it is often considered as such; it is exempt from income tax, and capital gains tax, much like other retirement savings scheme in other countries. Participants were presented with adapted versions of the ISA to reflect hypothetical descriptions of three different type of enrolment system: a mandated choice, default opt-in, or default opt-out. In Experiment "b", participants were HK residents in employment and were Cantonese speakers; this meant that the materials were prepared in Chinese; and were

translated by a Senior Researcher at Hang Seng Management School in HK, and back-translated to check for consistency in the UK and HK materials. The three descriptions of the enrolment systems were also adapted to the Mandatory Provident Fund (MPF – employer automatically enroll employees into a pension with savings being defaulted into pre-specified funds) for the HK sample. For full details of questionnaires and the specific instructions, please see Appendix B.

1.6. Participants and Design

Experiment 1a (UK sample). A total of $N = 155$ participants took part in an online survey (83 women; aged 18-77 years, $M_{age} = 45.12$ years, $SD = 15.10$ years). They were all UK residents recruited from the Qualtrics panel, an online platform for running experiments. A fixed payment was received by each participant as an incentive (10.00 GBP approximately 12.97 USD at the time of the study). Participants were randomly allocated to one of the three enrolment systems: Default opt-in ($n = 53$), default opt-out ($n = 51$), and mandated choice ($n = 51$). The dependent variable here is the general assessment of people's preference to register, perceived belief of an investor's true preference to register, perceived diagnosticity of people's intention to register in the retirement savings scheme measured by Bayesian Likelihood Ratio, and likelihood estimates of participants registering in a retirement savings scheme under the same enrolment system as Mark.

Experiment 1b (HK sample). A total of $N = 155$ participants took part in an online survey (76 females; aged 18-63 years, $M_{age} = 32.86$ years, $SD = 8.90$ years). They were all HK residents recruited from the Qualtrics panel, an online platform for running experiments. A fixed payment was received by each participant as an incentive (10.00 GBP approximately 12.97 USD at the time of the study). Participants were randomly allocated to one of the three enrolment systems: Default opt-in ($n = 51$), default opt-out ($n = 51$), and mandated choice ($n = 53$). The dependent variables were identical to Experiment 1a.

Experiment 2a (UK sample). A total of $N = 155$ participants took part in an online survey (71 women, 84 men; aged 25-65 years, $M_{age} = 44.45$ years, $SD = 10.53$ years). They were all UK residents recruited from the Qualtrics panel, an online platform for running experiments. A fixed payment was received by each participant as an incentive (10.00 GBP approximately 12.97 USD at the time of the study). An independent design was used in which participants

were randomly allocated to one of the three enrolment systems: Default opt-in ($n = 52$), default opt-out ($n = 51$), and mandated choice ($n = 52$). The dependent variables are the effect of enrolment systems on the general assessment of people's preference to register, perceived belief of an investor's true preference to register, and perceived diagnosticity of an investor's intention to register in a retirement savings scheme under the different enrolment systems.

Experiment 2b (HK sample). A total of $N = 150$ participants took part in an online survey (86 women, 74 men; aged 18-64 years, $M_{age} = 36.91$ years, $SD = 9.78$ years). They were all HK residents recruited from the Qualtrics panel, an online platform for running experiments. A fixed payment was received by each participant as an incentive (100 HKD approximately equivalent to 12.89 USD at the time of the study). An independent design was used in which participants were randomly allocated to one of the three enrolment systems: Default opt-in ($n = 52$), default opt-out ($n = 41$), and mandated choice ($n = 57$). The dependent variables are identical to Experiment 2a.

1.7. General Materials and Procedure

Experiment 1a (UK sample). Participants read a brief description of the ISA scheme and were asked to answer Question 1 to elicit their prior belief, $P(h)$: "Out of 100 people who live in the UK, for how many do you think their true preference is to be registered in the ISA scheme?", on a scale of 0 to 100. This was designed to investigate participants' general assessment of people's preference to register for the ISA scheme. Next, participants read details about a person named Mark living in an area that falls under a hypothetical enrolment system (they were assigned to one of the experimental conditions which reflected the three different enrolment types: Default opt-in, default opt-out, and mandated choice – see Figure 1). Under each enrolment system, participants were told that Mark is registered in the ISA scheme. Based on the definition of the enrolment system, participants were required to answer the following questions in the following order: Question 2 elicited their perceived belief about an investor's (i.e., Mark's) true preference to register for the ISA scheme: "How likely do you think it is that Mark's true preference was to be registered in the ISA default scheme?", on a scale of 0 (Mark definitely did NOT want to be registered in the ISA default scheme) to 100 (Mark definitely

DID want to be registered in the ISA default scheme).³ Question 3 asked “If we assembled 100 people whose true preference is to be registered in an ISA scheme, how many of them do you think will be registered in the ISA scheme on ... system?”, and Question 4 asked “If we assembled 100 people whose true preference is NOT TO be registered in an ISA scheme, how many of them do you think will be registered in the ISA scheme in the ... system?”, on a scale of 0 to 100, where “...” represents the enrolment system they were assigned to. Question 3 and

Question 4 were designed to elicit $\frac{P(\text{registration}|\text{want to invest})}{P(\text{registration}|\text{DON'T want to invest})}$ for calculating the likelihood ratio. Finally, participants were asked to imagine that they now live in the same city as Mark (in which ... system is in place). Based on this, they were asked Question 5: “How likely do you think that you will be registered in the ISA default scheme?”, on a scale of 0 (very unlikely) to 100 (very likely), to assess the likelihood estimates of participants registering in the same retirement savings scheme as Mark.

Experiment 1b (HK sample). The design and materials were the same as Experiment 1a with the exception that the details regarding the retirement savings scheme were changed from ISA to MPF (Figure 1) in order to reflect the current retirement savings scheme available in HK.

³ A frequency format was adopted to elicit the relevant Bayesian parameters as this has been shown to be an easier way for participants to provide responses (see also, e.g., Gigerenzer & Hoffrage (1995)).

UK Sample	HK Sample
<p>Default Opt-In Condition</p> <p>Mark lives in an area in the City with an OPT IN banking system.</p> <p>Under this system, everyone is automatically registered as NON-ISA Default, meaning that they will NOT have their money automatically invested by their bank/building society in the ISA SCHEME.</p>	<p>Default Opt-In Condition</p> <p>Mark lives in an area in the City with an OPT IN banking system.</p> <p>Under this system, everyone is automatically registered as NON-MPF Default, meaning they will NOT have their money automatically invested by their bank/insurer in the MPF SCHEME.</p>
<p>Default Opt-Out Condition</p> <p>Mark lives in an area in the City with an OPT OUT banking system.</p> <p>Under this system, everyone is automatically registered as ISA default, meaning that they will automatically have their money invested for them by their bank/building Society.</p>	<p>Default Opt-Out Condition</p> <p>Mark lives in an area in the City with an OPT OUT banking system.</p> <p>Under this system, everyone is automatically registered as MPF default, meaning that they will automatically have their money invested for them by their Bank/Insurer.</p>
<p>Mandated Choice Condition</p> <p>Mark lives in an area in the City with a CHOICE banking system.</p> <p>Under this system, everyone is legally required by their bank/building society to choose between being a NON-ISA-Default MEMBER or a ISA-Default MEMEBER.</p>	<p>Mandated Choice Condition</p> <p>Mark lives in an area in the City with a CHOICE banking system.</p> <p>Under this system, everyone is legally required by their bank/insurer to choose between being a NON-MPF-Default MEMBER or a MPF-Default MEMEBER.</p>

Figure 1. Description of the enrolment systems in each experimental condition for both the UK sample and HK sample.

Experiment 2 aimed to replicate the main results from Experiment 1. The experimental set up of Experiment 2 was the same as Experiment 1 with the exception of the prior belief question being asked after the presentation of the different enrolment systems. In addition, Questions 5 in Experiment 1 which elicited the likelihood of participants registering in the same retirement savings scheme as Mark was excluded in Experiment 2. Lastly, wording of the prior belief question and likelihood ratio question was modified to better reflect the cover story (see more details in below).

Experiment 2a (UK sample). Participants read a brief description of the ISA scheme and details about a person named Mark living in an area that falls under a hypothetical enrolment system (they were assigned to one of the experimental conditions which reflected the three different enrolment types: Default opt-in, default opt-out, and mandated choice – see Figure 1). Under each enrolment system, participants were told that Mark is registered in the ISA scheme. Based on the definition of the enrolment system, participants were required to answer the following questions in the following order: Question 1 elicited their perceived belief about an investor’s (i.e., Mark’s) true preference to register for the ISA scheme: “*How likely do you think it is that Mark’s true preference was to be registered in the ISA default scheme?*”, on a scale of 0 (Mark definitely did NOT want to be registered in the ISA default scheme) to 100 (Mark definitely DID want to be registered in the ISA default scheme); Question 2 elicited

their *prior belief*, $P(h)$: “Out of 100 who live in the same area as Mark, for how many do you think their true preference is to be registered in the ISA default scheme?”, on a scale of 0 to 100. This was designed to investigate participants’ general assessment of people’s preference to register for the ISA scheme. Question 3 asked “If we assembled 100 people whose true preference is to be a member of the ISA default scheme, how many of them do you think will be registered in the ISA default scheme on ... system?”, and Question 4 asked “If we assembled 100 people whose true preference is NOT TO be a member of the ISA default scheme, how many of them do you think will be registered in the ISA default scheme on the ... system?”, on a scale of 0 to 100. Question 3 and Question 4 were elicited to calculate the likelihood ratio

$$\frac{P(\text{registration}|\text{want to invest})}{P(\text{registration}|\text{DON'T want to invest})}$$

Experiment 2b (HK sample). The design and materials were the same as Experiment 2a with the exception that the details regarding the retirement savings scheme were changed from ISA to MPF in order to reflect the current retirement savings scheme available in HK.

1.8.Results

Experiment 1a and 1b: General assessment of people’s preference to register, $P(h)$

UK participants’ general assessment of people’s preference to register, $P(h)$, in the ISA scheme [$M = 44.46$, $SD = 19.58$] was significantly higher than HK participants’ $P(h)$ in the MPF scheme [$M = 37.51$, $SD = 26.13$], $t(308) = 2.65$, $p = .008$, $d = .30$, 95% CI[1.79, 12.12].

Experiment 2a and 2b: Effect of enrolment systems on $P(h)$

In both samples, participants who read about the different enrolment systems (Experiment 2) were more likely to give higher general assessment of people’s preference to register, $P(h)$, compared to Experiment 1 (i.e., $P(h)$ elicited prior to reading about the different enrolment systems), $t(613) = 5.26$, $p < .001$, $d = .42$, 95% CI [6.28, 13.76]; but there was no difference in $P(h)$ between the three different enrolment systems in Experiment 2a, $F(2, 152) = 2.79$, $p = .065$, $\eta^2 = .04$; and in Experiment 2b, $F(2, 147) = 2.49$, $p = .086$, $\eta^2 = .03$. Nonetheless, consistent with Experiment 1, UK participants’ $P(h)$ in the ISA scheme [$M = 49.57$, $SD = 21.82$] were significantly higher than HK participants’ $P(h)$ in the MPF scheme [$M = 42.29$, $SD = 25.77$], $t(613) = 3.78$, $p < .001$, $d = .30$, 95% CI [3.50, 11.06].

Perceived belief of an investor's true preference to register

Experiment 1a (UK sample). The findings are broadly consistent with our prediction such that the perceived belief of an investor's (i.e., Mark's) true preference to register in the ISA scheme reflected sensitivity to the three different enrolment systems, and decreased in the following order: Mandated choice = Default opt-in > Default opt-out (Figure 2). A one-way ANOVA revealed that the enrolment systems had a significant effect on participants' perceived belief of an investor's (i.e., Mark's) true preference to register for the ISA scheme, $F(2, 152) = 6.81, p = .001, \eta^2 = .08$. The default opt-in system was judged as a stronger indicator of true preference to register for the ISA scheme when compared with the default opt-out system, $t(152) = 3.40, p = .001, d = .65, 95\% \text{ CI } [4.96, 27.77]$. The mandated choice system was judged as a stronger indicator of true preference when compared with the default opt-out system, $t(152) = 2.96, p = .004, d = .62, 95\% \text{ CI } [2.88, 25.90]$. There was no significant difference between the default opt-in system and mandated choice system, $t(152) = .410, p = .683, d = .08, 95\% \text{ CI } [-9.43, 13.38]$.

Experiment 1b (HK sample). The findings were broadly consistent with our prediction that the perceived belief of an investor's true preference to register for the MPF scheme reflected sensitivity to the three different enrolment systems (with the exception of mandated choice), and decreased in the following order: Default opt-in > Mandated choice = Default opt-out (Figure 2). A one-way ANOVA revealed a significant effect of the enrolment systems on participants' perceived belief of an investor's (i.e., Mark's) true preference to register in the MPF scheme, $F(2, 152) = 6.53, p = .002, \eta^2 = .08$. Planned pairwise comparison show that the default opt-in system was judged as a stronger indicator of true preference when compared with the default opt-out system, $t(152) = 3.52, p = .001, d = .71, 95\% \text{ CI } [6.18, 31.62]$, and the mandated choice system, $t(152) = 2.49, p = .014, d = .46, 95\% \text{ CI } [.65, 25.85]$. There was no significant difference between the default opt-out system and the mandated choice systems, $t(152) = 1.06, p = .290, d = .22, 95\% \text{ CI } [-6.95, 18.25]$.

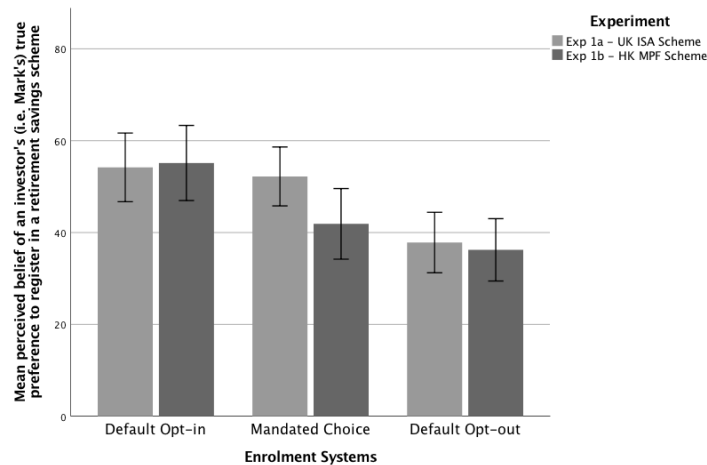


Figure 2. Mean perceived belief of an investor's (i.e. Mark's) true preference to register in the ISA scheme (Experiment 1a, UK sample) and in the MPF scheme (Experiment 1b, HK sample) across the three experimental conditions. Error bars are 95% CI.

Experiment 2a (UK sample). The findings supported our general prediction that the perceived belief of an investor's true preference to register in the ISA scheme reflected sensitivity to the three different enrolment systems, and decreased in the following order: Mandated choice = Default opt-in > Default opt-out (Figure 3). This is consistent with the patterns found in Experiment 1a. The evidence revealed that the enrolment systems had a significant effect on participants' perceived belief of an investor's (i.e., Mark's) true preference to register in the ISA scheme, $F(2, 152) = 11.95, p < .001, \eta^2 = .14$. Planned pairwise comparisons confirmed the trend indicated in Figure 4. The default opt-in system was judged as a stronger indicator of true preference when compared with the default opt-out system, $t(152) = 4.32, p < .001, d = .83, 95\% \text{ CI } [9.34, 32.04]$. The mandated choice system was judged as a stronger indicator of true preference when compared with the default opt-out system, $t(152) = 4.16, p < .001, d = .86, 95\% \text{ CI } [8.61, 31.31]$. There was no significant difference between the default opt-in system and the mandated choice system, $t(152) = .15, p = .878, d = .03, 95\% \text{ CI } [-10.56, 12.02]$.

Experiment 2b (HK sample). The pattern of findings from Experiment 2b was broadly similar to those reported in Experiment 1b (and also in line with our general prediction). The default opt-in system was perceived as more indicative of an investor's true preferences, than the default opt-out system: Default opt-in > Default opt-out, Default opt-in = Mandated choice, Default Opt-out = Mandated choice (Figure 3). The evidence revealed that the enrolment system had a significant effect on participants' judgments of Mark's true preference, $F(2, 147)$

= 3.26, $p = .041$, $\eta^2 = .04$. Planned pairwise comparisons revealed that the default opt-in system was judged as a stronger indicator of an investor's (i.e., Mark's) true preference when compared with the default opt-out system, $t(147) = 2.54$, $p = .012$, $d = .53$, 95% CI [.93, 26.74]. However, there was no significant difference between the default opt-in system and the mandated choice system, $t(147) = .96$, $p = .337$, $d = .18$, 95% CI [-7.04, 16.67], and between the default opt-out system and the mandated choice system, $t(147) = 1.69$, $p = .094$, $d = .36$, 95% CI [-3.64, 21.67].

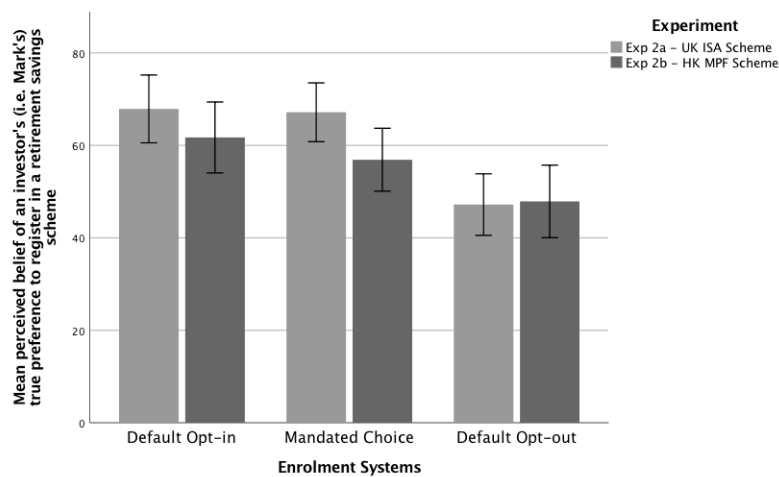


Figure 3. Mean perceived belief of an investor's (i.e. Mark's) true preference to register in the ISA scheme (Experiment 2a, UK sample) and in the MPF scheme (Experiment 2b, HK sample) across the three experimental conditions. Error bars are 95% CI.

Perceived diagnosticity of people's intention to register

A one-way ANOVA was conducted to analyse perceived diagnosticity of evidence inferred

from the likelihood ratio $\frac{P(\text{registration}|\text{want to invest})}{P(\text{registration}|\text{DON'T want to invest})}$.

Experiment 1a (UK sample) and Experiment 1b (HK sample). The analysis revealed a non-significant effect of enrolment system on the perceived diagnosticity of people's intention to register in the ISA scheme, $F(2, 143) = .622$, $p = .538$, $\eta^2 = .01$ (default opt-in [$M = 3.35$, $SD = 7.53$], default opt-out [$M = 2.08$, $SD = 3.69$], and mandated choice [$M = 2.73$, $SD = 5.26$]). Similarly, there was also a non-significant effect of enrolment system on the perceived diagnosticity of people's intention to register in the MPF scheme, $F(2, 115) = .542$, $p = .583$,

$\eta^2 = .01$ (Default opt-in [$M = 3.26, SD = 3.91$], default opt-out [$M = 4.49, SD = 14.74$], and mandated choice [$M = 2.35, SD = 2.54$]).

Experiment 2a (UK sample) and Experiment 2b (HK sample). Consistent with Experiment 1, the analysis revealed a non-significant effect of enrolment system on the perceived diagnosticity of people's intention to register for the ISA scheme, $F(2, 137) = 1.40, p = .249, \eta^2 = .02$ (Default opt-in [$M = 2.99, SD = 7.50$], default opt-out [$M = 1.32, SD = .91$], and mandated choice [$M = 2.09, SD = 3.36$]). Similarly, there was also a non-significant effect of enrolment system on the perceived diagnosticity of people's intention to register for the MPF scheme, $F(2, 132) = 1.71, p = .185, \eta^2 = .03$ (Default opt-in [$M = 5.98, SD = 14.36$], default opt-out [$M = 2.64, SD = 7.82$], and mandated choice [$M = 2.75, SD = 4.07$]).

Likelihood estimates of participants registering in a retirement savings scheme under the same enrolment system as Mark⁴

Experiment 1a (UK Sample). When asked to imagine that they lived in the same city as Mark, with the same enrolment system implemented, a one-way ANOVA revealed a non-significant effect of enrolment systems on the likelihood estimates of participants registering in the ISA scheme, $F(2, 152) = 2.23, p = .11, \eta^2 = .03$.

Experiment 1b (HK Sample). A one-way ANOVA revealed a significant effect of enrolment systems on the likelihood estimates of participants registering in the MPF scheme, $F(2, 152) = 3.16, p = .045, \eta^2 = .04$ (Figure 4). Planned pairwise comparison show that participants were significantly more likely to register for the MPF in the default opt-out system than the mandated choice system, $t(152) = 2.48, p = .014, d = .20, 95\% CI [.65, 27.92]$. There was no significant difference between the default opt-in system and the mandated choice system, $t(152) = .85, p = .395, d = .28, 95\% CI [-8.72, 18.54]$; and between the default opt-out system and default opt-in system, $t(152) = 1.61, p = .109, d = .11, 95\% CI [-4.39, 23.14]$.

⁴ This question was only asked for Experiment 1a and 1b.

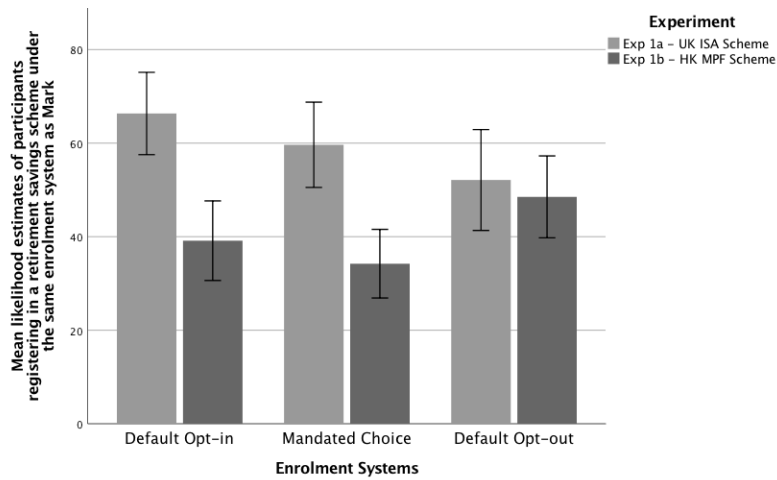


Figure 4. The likelihood estimates of participants registering in a retirement savings scheme under the same enrolment system as Mark across the three experimental conditions in Experiment 1a (UK Sample) and Experiment 1b (HK Sample). Error bars are 95% CI.

1.9. Discussion

The focus of the present study was to examine part one of the second objective of the thesis – the welfarist argument – under which nudge undermines autonomy. To do so, the study considered judgments regarding whether nudges capture people’s true preferences to register in retirement savings schemes. This is of particular importance given the complexity of the domain in which people are required to make decisions, and the limited knowledge people generally have around financial products (Anthes, 2004; Bucher-Koenen & Lusardi, 2011; Lusardi & Mitchell, 2011a, 2011b; Song, 2015; Van Rooij et al., 2011). This study investigated judgments on the perceived belief of an investor’s true preference across three enrolment systems (Default opt-in, Default opt-out, Mandated choice) using samples which have experience of different enrolment systems in their country (i.e., free choice system in the UK, mandated system in HK).

Experiments 1a and 1b. The findings (both the UK and HK samples) supported the main hypothesis which is that when a hypothetical investor has been ‘defaulted’ into a retirement savings scheme under a default opt-out system, their true preference to register is perceived as weaker than when the choice was made actively (under the mandated choice and default opt-in system); with the exception of Experiment 1b where there was no significant difference between the mandated choice system and the default opt-out system. Subsequently, when asked to imagine that they live in the same city as Mark with the same enrolment system

implemented, HK participants were less likely to register in the MPF scheme under the mandated choice system compared to the default opt-out system. Based on the analysis of the general assessment of people's preference to register, $P(h)$, the suspected difference observed between the samples reflects their prior experience of enrolment systems in their respective countries, as HK participants generally gave lower $P(h)$ compared to their UK counterparts. Lastly, the perceived diagnosticity of evidence inferred from the likelihood ratio did not reveal a significant difference between the three enrolment systems.

Experiment 2a and 2b. Given the differences in the pattern of findings are potentially influenced by the different prior experiences between the samples, Experiment 2a and 2b were conducted to replicate the main behavioural results in Experiment 1a and 1b and explore the extent to which $P(h)$ is sensitive to the framing of the three different enrolment systems. In line with the main hypothesis, the pattern of results obtained from Experiments 2a and 2b suggests that the enrolment system impacted people's perceived belief of an investor's (i.e., Mark's) true preference to register in a retirement savings scheme. More specifically, when an investor has been 'defaulted' into a retirement savings scheme under a default opt-out system their perceived true preference for that option was judged as weaker than when the choice was made actively under the mandated choice or default opt-in systems (with the exception of Experiment 2b whereby the mandated choice system was not judged significantly different from the default opt-out system – this is consistent with Experiment 1b).

These observed differences between the samples can be explained by the prior experiences of the different enrolment systems in the samples' respective countries. Consistent with Experiment 1, HK participants' general assessment of people's preference to register, $P(h)$, were judged as weaker than those of UK participants. In addition, introducing the enrolment systems lead to overall higher $P(h)$ compared to those elicited in Experiment 1, but there were no differences in $P(h)$ between the three different enrolment systems. This may suggest that the introduction of the enrolment systems did not influence participants prior experience of different enrolment systems in their respective countries whilst their perceived belief of an investor's true preference to register was based on their prior experiences. Lastly, also consistent with Experiment 1, the perceived diagnosticity of evidence inferred from the likelihood ratio did not reveal significant differences between the three enrolment systems.

Overall, the findings supported the hypothesis that people are sensitive to the different framings of the decision to be a registered investor in the enrolment system and they, therefore, see registration differentially representative of an investor's true preference to register. In two experiments, the general pattern of findings within the respective samples was similar. For the UK sample (Experiment 1a and 2a), there was no difference between the mandated choice system and the default opt-in system in participants' perceived belief of an investor's true preference to register in the ISA scheme, but both were judged as more indicative of an investor's true preference than the default opt-out system. For the HK sample (Experiments 1b and 2b), there was no difference between the mandated choice system and the default opt-out system, but both were judged as a weaker indicator of an investor's true preference to register in the MPF scheme than the default opt-in system (with the exception of Experiment 2b, where no differences were found between the default opt-in system and the mandated choice system). Across both samples, and across both experiments, the default opt-out system was generally judged as the weakest indicator of the investor's true preference to register for a retirement savings scheme. This finding satisfies the welfarist argument, because a lack of autonomy to make one's own decisions is perceived as a weaker indicator of one's true preference. I will discuss the implication of these findings in Chapter 7.

CHAPTER 5: Underlying Wishes and Nudged Choices in Pro-social Context

Chapter 3.1 examined part one of the second objective of the thesis – the welfarist argument – under which nudge undermines autonomy in pro-self and pro-social contexts. In this Chapter, I will provide the empirical evidence for the welfare consequence of implementing pro-social defaults in the organ donation context. The findings from Chapter 4 (pro-self context) revealed that people’s judgments of an individual’s true preference were sensitive to the framing of enrolment systems. The default opt-in system was generally perceived as a stronger indicator of an individual’s true preference to register in a retirement savings scheme than the default opt-out system. The main focus of this chapter will be on providing empirical findings on people’s judgement of individual’s “true” preference to register in the organ donation systems in a pro-social context, based also on the application of a Bayesian approach, and to see if the main findings replicate those in the retirement savings context.

The present study examines third-party judgements of people’s choices made under different choice systems (default and non-default). This is because decisions made under a default opt-out system (a passive choice) is perceived as more ambiguous than decisions under a default opt-in or mandated choice system (an active choice). This has implication on family consent rate because they have the opportunity to consent or veto organ donation on behalf of their relative depending on what signals are being perceived from the organ donation systems. This issue is consequential in the context of increasing actual donation rates.

In line with the methodology adopted in Chapter 4, this study investigates people’s perceptions of an individual’s true preference from their registration on an organ donation register, here we are interested in the perceived diagnosticity of this evidence for inferring their true preference. From a Bayesian perspective, the diagnosticity of an item of evidence, registration on the ODR, is determined by the likelihood ratio. The likelihood ratio is the ratio of obtaining that evidence in the event that the underlying hypothesis is true (the relative’s underlying wishes were to donate their organs - ‘hit’ rate) as opposed to obtaining that evidence in the event that the underlying hypothesis is false (the relative’s underlying wishes were to *not* donate their organs - ‘false positive’ rate), $\frac{P(\text{registration}|\text{want to donate})}{P(\text{registration}|\text{DON'T want to donate})}$.

This study included an additional independent variable “mandatory donor” to ensure that people are giving proper judgements because if people are providing sensible/genuine judgements then this option should be the least preferred; thus, acting as a type of control. The present study also included additional demographic data to provide further insights on the influence of individual donor status on the dependent measures.

The specific hypotheses investigated here are:

H1(pro-social welfare). People recognize the consequences of the different framings of the decision to be a registered organ donor. They, therefore, see registration as differentially representative of an individual’s true preference to donate, in the following order: Default opt-in > Mandated choice > Default opt-out > Mandatory Donor.

H2 (pro-social welfare). If Hypothesis 1 is supported, then families making decisions on the basis of their beliefs about the deceased’s true preferences to donate will be perceived by participants as more likely to agree to donation in line with the predictions of Hypothesis 1.

Davidai et al. (2013) found that the meaning individuals attach to participating in organ donation was different between default opt-out and default opt-in countries. More specifically, the willingness to donate one’s organs in an opt-in country was considered roughly akin to giving away one’s wealth to charity upon one’s death, and in an opt-out country, this fell between letting other get ahead of one in line and volunteering some time to help the poor. Their study showed that when asked explicitly to consider the meaning of the relevant choices, people are capable of discerning this difference in meaning. Therefore, to eliminate this as a factor in the study, the current study investigated third party’s judgement of an individual’s true preference to donate in both opt-in and opt-out countries. In general, the materials presented in Experiment 1-3 were the same, except for some additional questions in Studies 2 and 3. These additional questions were based on the manipulations regarding the samples used: Experiment 1 involved US citizens (the US has a default opt-in system), Experiment 2 involved citizens from European countries that has a default opt-in system, and Experiment 3 involved citizens from European countries that has a default opt-out system. It was essential to present a set of studies that systematically address the main hypotheses, therefore as few details as possible were carefully changed between the studies. Given the similarities between these studies, and to assist the reader, methods and results of the three studies are presented together

to facilitate a better understanding of the overall pattern of findings. For all three experiments ethics approval from QMUL college ethics board was granted, QMERC2014/54. For full details of questionnaires and specific instructions, please see Appendix C.

5.1. Participants and Design

Experiment 1. A total of $N = 493$ US citizens (Opt-in Policy) recruited from Amazon Mechanical Turk took part in the online survey (256 females; aged 18-72 years, $M_{\text{age}} = 35.65$ years, $SD = 11.65$ years), and were compensated \$0.50 for their participation.

Experiment 2. A total of $N = 401$ European citizens from countries with a default opt-in system (Germany, Denmark, Lithuania, the Netherlands, Romania, UK) were recruited from Prolific Academic (244 females; aged 18-72 years, $M_{\text{age}} = 35.50$ years, $SD = 12.76$ years), and were compensated \$0.50 for their participation.

Experiment 3. A total of $N = 400$ European citizens from countries with a default opt-out system (Austria, Spain, France, Italy, Belgium, Sweden, Greece, Finland, Poland, Portugal, Turkey, Slovenia, Croatia, Czech Republic, and Norway) were recruited from Prolific Academic (261 females; aged 18-60 years, $M_{\text{age}} = 29.01$ years, $SD = 8.27$ years), and were compensated \$0.50 for their participation.

5.2. General Materials and Procedure

In each experiment, participants were randomly allocated to one of four experimental conditions (Figure 5):

Experiment 1: Default opt-in ($n = 122$), mandated choice ($n = 128$), default opt-out ($n = 118$), and mandatory donor ($n = 125$) with four dependent variables (Perceived belief of the donor's true preference to donate; Perceived likelihood estimates of a relative consenting to donate; Relative's consent expressed as a binary choice; Perceived diagnosticity of people's intention to donate measured by Bayesian Likelihood Ratio).

Experiment 2: Default opt-in ($n = 101$), mandated choice ($n = 100$), default opt-out ($n = 101$), mandatory donor ($n = 99$). In addition to the dependent variables in Experiment 1 (except for relative's consent expressed as a binary choice which was replaced with dropdown list reasons

for relatives consenting or vetoing donation), we also included two other questions (these are presented in more detail below), and minor modifications to the elicitation of prior belief question.

Experiment 3: Default opt-in (n = 99), mandated choice (n = 100), default opt-out (n = 99), mandatory donor (n = 102). The dependent variables were identical to those included in Experiment 2.

Default Opt-In System

Mark lives in an area with an **OPT IN** system of organ donation.
Under this system, everyone is automatically registered as a **NON-DONOR**, meaning they will **NOT** have their organs used in the event of their death. Anyone who wishes to be a donor must make an extra effort by going online and changing their preferences or by calling the donor line.

Default Opt-Out System

Mark lives in an area with an **OPT OUT** system of organ donation.
Under this system, everyone is automatically registered as a **DONOR**, meaning they will have their organs used in the event of their death. Anyone who wishes to be a **NON-DONOR** must make an extra effort by going online and changing their preferences or by calling the donor line.

Mandated Choice System

Mark lives in an area with a **CHOICE** system of organ donation.
Under this system, everyone is legally required to choose between being a **DONOR** OR **NON-DONOR** before they register for their driver's license.

Mandatory Donor System

Mark lives in an area with a **DONOR** system of organ donation.
Under this system, everyone is automatically **REGISTERED AS A DONOR**. There is no option for changing this.

Figure 5. The scenario described in each organ donation legislative systems: Default Opt-In, Default Opt-out, Mandated Choice and Mandatory Donor.

Experiment 1. Participants were presented with an online questionnaire. They were told that this was a social experiment designed to investigate the topic of organ donation. The questionnaire first provided the definition of organ donation and, based on this information, participants were required to answer a question to elicit prior beliefs about people's true preferences to donate, 1) *P(want to donate)*: "Out of 100 people in the U.S., for how many do

*you think their true preference is to donate their organs?”*⁵ A fictional scenario immediately followed in which participants read details about a person named Mark who lives in an area that falls under a particular organ donation legislative system (depending on the experimental condition participants were randomly assigned to one of the four organ donation legislative systems: default opt-in, default opt-out, mandated choice, or mandatory donor – see Figure 5).

In all four conditions participants were explicitly told that Mark was registered as a donor. They were also told that Mark was involved in a fatal accident leaving his vital organs intact. Following on from this, participants were asked 2) *“How likely do you think it is that Mark’s true preference was to donate his organs?”* to elicit perceived belief of true preference to donate, on a scale of 1 (Mark definitely did not want to donate his organs) to 100 (Mark definitely did want to donate his organs). In addition, two other questions were presented to assess what actions should follow given the news that Mark suffered a fatal accident. Participants were asked 3) *“How likely is it that John will agree to his Uncle’s (Mark) organs being donated?”*; responses were provided on a scale from 1 (very unlikely) to 5 (very likely) to elicit perceived likelihood estimates of a relative consenting to the donor’s decision to donate. Finally, participants were asked 4) *“What will John decide to do, will he donate his Uncle’s organs?”* so as to elicit a binary response (consent, veto) as to the relative’s decision.

Questions 5 and 6 elicited the relevant conditional probabilities for calculating the likelihood ratio, which is a measure of the perceived diagnosticity of the deceased’s registration on the ODR. Participants were asked 5) *“If we assembled 100 people whose true preference is to DONATE their organs, how many of them do you think will be registered as organ donors on the system?”*. And then asked 6) *“If we assembled 100 people whose true preference is NOT TO DONATE their organs, how many of them do you think will be registered as organ donors on the system?”*, on a scale of 0-100 (where “...” is the organ donation legislative system participants were assigned to).

Experiment 2 and Experiment 3. The main set-up is the same as in Experiment 1 but with minor modifications. First, questions that had the term “true preference” were replaced with

⁵ A frequency format was adopted to elicit the relevant Bayesian parameters as this has been shown to be an easier way for participants to provide responses (see also, e.g., Gigerenzer & Hoffrage (1995)).

the word “want” so that the question is better understood. Second, the cover story now describes the respective organ donation legislative systems before eliciting participant’s prior belief, Question 1) “*Out of 100 people living in this country, for how many do you think would want to donate their organs?*”. Followed by Question 2) “*How likely do you think it is that Mark wanted to donate his organs?*”, the scale has now explicitly labelled the midpoint which ranges from 0 (Mark definitely did not want to donate his organs, 50 (Mark is equally likely to donate or not donate his organs), to 100 (Mark definitely did want to donate his organs). After responding to questions 1 and 2, participants were then told about John and were asked to provide ratings to two additional new questions 3) “*To what extent does being registered to donate under the ... system provide a clear indication that Mark wanted to donate his organs?*”, on a scale of 0 (Not at all clear) to 100 (Absolutely clear); this question was designed to elicit perceived signal of intent to donate, and 4) “*How likely it is that John believes that Mark wanted to donate his organs?*”, on a scale of 0 (John believes Mark definitely did not want to donate his organs), 50 (John thinks it is equally likely that Mark wanted to donate his organs as didn’t), to 100 (John believes that Mark definitely wants to donate his organs); this question was designed to elicit the likelihood estimates of relative’s belief of the donor’s true preferences to donate.

Depending on the response to Question 5) “*How likely it is that John will agree to his Uncle’s organs being donated?*”, participants were presented with additional new question 6) “*Why do you think it is highly unlikely/moderately unlikely/likely/moderately likely/highly likely that John will donate Mark’s organs?*”, which replaced Question 4 (relative’s consent expressed as a binary choice) in Experiment 1. Participants were presented with up to eleven options as candidate reasons and an option for “none of the above” (five different reasons were presented for a “highly likely”, “moderately likely” and “likely” judgments, and eleven different reasons were presented for “highly unlikely” and “moderately unlikely” judgments) from a dropdown menu. The options presented to people were based on a prior study on potential causes for family refusal decisions (Ghorbani et al., 2011) and the most frequent reasons that emerged in practice (Vincent & Logan, 2012). The reasons participants could select from were: it is a highly traumatic time for relatives and it is just not something they can think about; lack of understanding of the organ donation process; denial and rejection of brain-death criteria; the hope for a miracle; fear about organ donation trade and unknown organ destination; religious beliefs; insecurity about the brain-death diagnosis; unsure about the deceased’s wish to donate;

belief in body integrity about death; fear of objection by other family members; and there is lack of evidence to indicate that the deceased wanted to donate; but note that this is by no means an exhaustive list. Finally, participants were presented with two questions to elicit the likelihood ratio, Question 7) “*If we assembled 100 people who want to DONATE their organs, how many of them do you think will be registered as organ donors on the ... system?*” and Question 8) “*If we assembled 100 people who DO NOT WANT to DONATE their organs, how many of them do you think will be registered as organ donors on the ... system?*”, on a scale of 0-100.

For all three studies, participants were asked a series of demographic questions (e.g., age, gender, religion), questions regarding their own organ donation status and their view on organ donation (e.g. whether they agree with a default opt-out system, willingness to agree to donation if loved one’s wishes are unknown, who should decide donation in the event of death, etc.)

5.3.Results

Perceived belief of the donor’s true preference to donate

The findings were broadly in line with our first hypothesis such that the participant’s perceived belief of the donor’s (i.e. Mark’s) true preference to donate reflected sensitivity to the four different organ donation legislative systems (Figure 6). Overall, the evidence revealed that the organ donation legislative system had a significant effect on participants’ perceived belief of the donor’s perceived true preference to donate [*Experiment 1: $F(3, 492) = 74.69, p < .001, \eta^2 = .31$; Experiment 2: $F(3, 400) = 41.55, p < .001, \eta^2 = .24$; Experiment 3: $F(3, 399) = 24.39, p < .001, \eta^2 = .16$].*

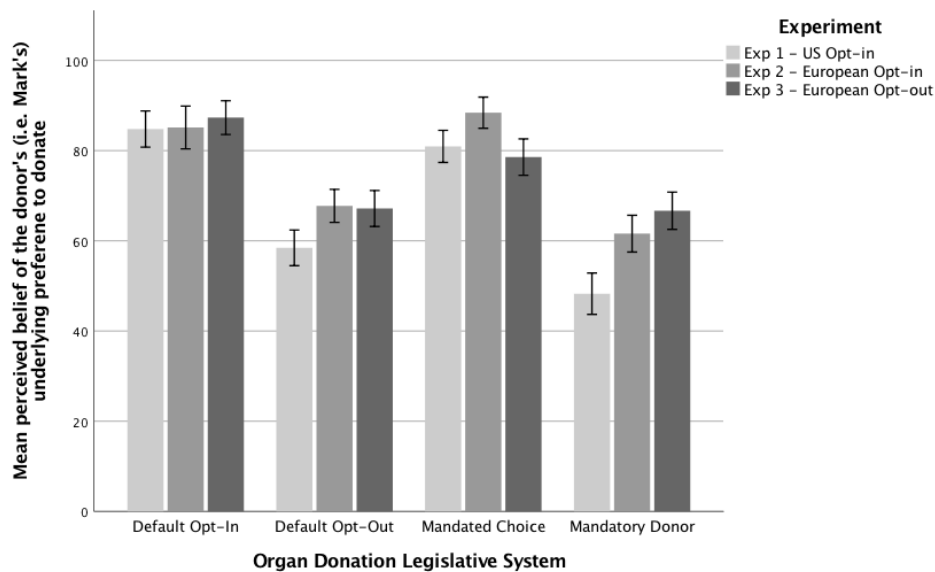


Figure 6. Mean perceived belief of the donor's (i.e. Mark's) true preference to donate across the four experimental conditions for Experiment 1, 2 and 3. Error bars are at 95% CI.

For Experiment 1 and 2, the participants' perceived belief of the donor's true preference to donate decreased in the following order: Mandated choice = Default opt-in > Default opt-out > Mandatory donor. Planned pairwise comparisons confirmed this pattern: The default opt-in system was perceived as a stronger indicator of true preference to donate when compared with the default opt-out system [*Experiment 1*: $t(489) = 9.00, p < .001, d = 1.20$; *Experiment 2*: $t(188) = 5.75, p < .001, d = 0.81$]; and the mandatory donor system [*Experiment 1*: $t(489) = 12.65, p < .001, d = 1.51$; *Experiment 2*: $t(194) = 7.47, p < .001, d = 1.06$]. The default opt-out system was judged as a stronger indicator of true preference to donate when compared with the mandatory donor system [*Experiment 1*: $t(489) = 3.49, p = .001, d = 0.43$; *Experiment 2*: $t(195) = 2.23, p = .027, d = 0.32$]. The mandated choice system was judged as a stronger indicator of true preference to donate when compared with the default opt-out system [*Experiment 1*: $t(489) = 7.78, p < .001, d = 1.07$; *Experiment 2*: $t(198) = 8.14, p < .001, d = 1.15$]; and the mandatory donor system [*Experiment 1*: $t(489) = 11.46, p < .001, d = 1.40$; *Experiment 2*: $t(192) = 9.96, p < .001, d = 1.41$]. There was no significant difference in true preference to donate between the default opt-in system and mandated choice system [*Experiment 1*: $t(489) = 1.34, p = .181, d = 0.18$; *Experiment 2*: $t(182) = 1.11, p = .269, d = 0.16$].

For Experiment 3, the participants' perceived belief of the donor's true preference to donate decreased in the following order: Default opt-in > Mandated choice > Default opt-out = Mandatory donor. Planned pairwise comparisons confirmed this pattern: The default opt-in system was judged as a stronger indicator of true preference to donate when compared with the default opt-out system [*Experiment 3*: $t(195) = 7.31, p < .001, d = 1.04$]; mandated choice system [*Experiment 3*: $t(196) = 3.16, p = .002, d = 0.45$]; and mandatory donor system [*Experiment 3*: $t(198) = 7.35, p < .001, d = 1.04$]. The mandated choice system was judged as a stronger indicator of true preference to donate when compared with the default opt-out system [*Experiment 3*: $t(197) = 3.99, p < .001, d = 0.57$]; and the mandatory donor system [*Experiment 3*: $t(200) = 4.06, p < .001, d = 0.58$]. There was no significant difference in true preference to donate between the default opt-out system and the mandatory donor system [*Experiment 3*: $t(199) = .17, p = .864, d = 0.02$].

Perceived signal of intent to donate

The findings were broadly in line with our first hypothesis. Participants' perceptions of the donor's (i.e. Mark's) intention to donate reflect sensitivity to the four different organ donation legislative systems (Figure 7); this question was not asked in *Study 1*, but was asked in *Experiment 2*: $F(3, 400) = 98.77, p < .001, \eta^2 = .43$, and *Experiment 3*: $F(3, 399) = 60.73, p < .001, \eta^2 = .32$.

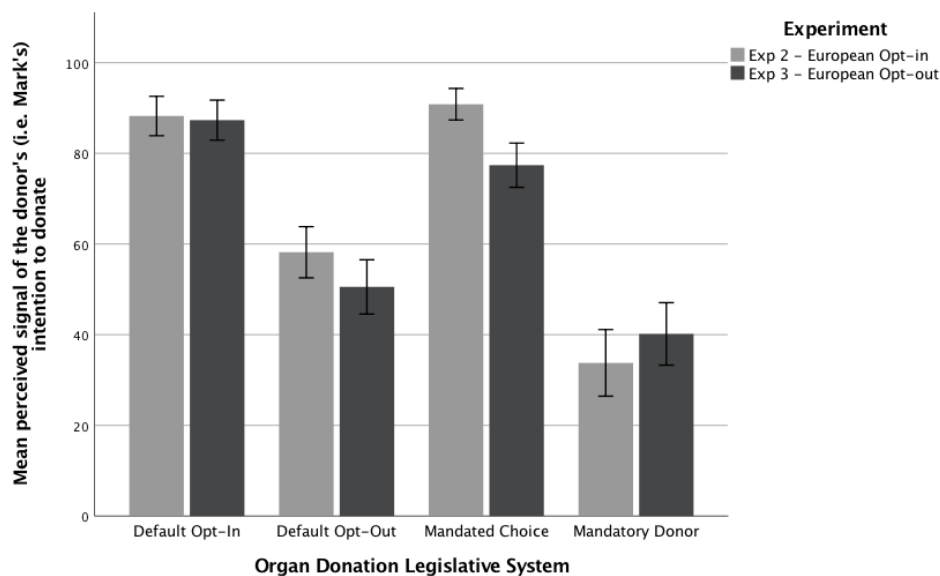


Figure 7. Mean perceived signal of the donor's (i.e. Mark's) intention to donate across the four experimental conditions for Experiment 2 and 3. Error bars are at 95% CI.

For Experiment 2, the participants' perceptions of the donor's intention to donate decreased in the following order: Mandated choice = Default opt-in > Default opt-out > Mandatory donor. This is consistent with the pattern found for perceived belief of the donor's true preference to donate. Planned pairwise comparisons confirmed this pattern: Default opt-in system was perceived as a stronger signal of intent to donate when compared with the default opt-out system [*Experiment 2*: $t(188) = 8.37, p < .001, d = 1.78$]; and the mandatory donor system [*Experiment 2*: $t(159) = 12.66, p < .001, d = 1.79$]. The mandated choice system was perceived as a stronger signal of intent to donate when compared with the default opt-out system [*Experiment 2*: $t(166) = 9.77, p < .001, d = 1.38$]; and the mandatory donor system [*Experiment 2*: $t(140) = 13.93, p < .001, d = 1.98$]. Similarly, the default opt-out system was perceived as a stronger signal of intent than the mandatory donor system [*Experiment 2*: $t(185) = 5.23, p < .001, d = 0.74$]. There was no significant difference for signal of intent to donate between the default opt-in system and the mandated choice system [*Experiment 2*: $t(190) = .93, p = .355, d = 0.13$].

For Experiment 3, the participants' perceptions of the donor's intention to donate decreased in the following order: Default opt-in > Mandated choice > Default opt-out > Mandatory donor. Planned pairwise comparisons confirmed this pattern: The default opt-in system was perceived as a stronger signal of intent to donate when compared with the default opt-out system [*Experiment 3*: $t(180) = 9.80, p < .001, d = 1.39$]; the mandated choice system [*Experiment 3*: $t(195) = 3.00, p = .003, d = 0.42$]; and the mandatory donor system [*Study 3*: $t(171) = 11.43, d = 1.61, p < .001$]. The mandated choice system was perceived as a stronger signal of intent to donate when compared with the default opt-out system [*Experiment 3*: $t(189) = 6.89, p < .001, d = 0.98$]; and the mandatory donor system, [*Experiment 3*: $t(181) = 8.74, p < .001, d = 1.23$]. Similarly, the default opt-out system was perceived as a stronger signal of intent to donate than the mandatory donor system [*Experiment 3*: $t(196) = 2.25, p = .025, d = 0.32$].

Perceived likelihood estimates of the relative's belief of the donor's true preference to donate

The findings were broadly in line with our second hypothesis such that the perceived likelihood estimates of the relative's belief of the donor's true preference to donate reflected sensitivity to the four different organ donation legislative systems (Figure 8). Overall, the evidence

revealed that the organ donation legislative system had a significant effect on participants' perceived likelihood estimates of the relative's belief of the donor's true preference to donate [this question was not asked for *Experiment 1*, but was asked in *Experiment 2*: $F(3, 400) = 64.23, p < .001, \eta^2 = .33$; and *Experiment 3*: $F(3, 399) = 29.84, p < .001, \eta^2 = .18$].

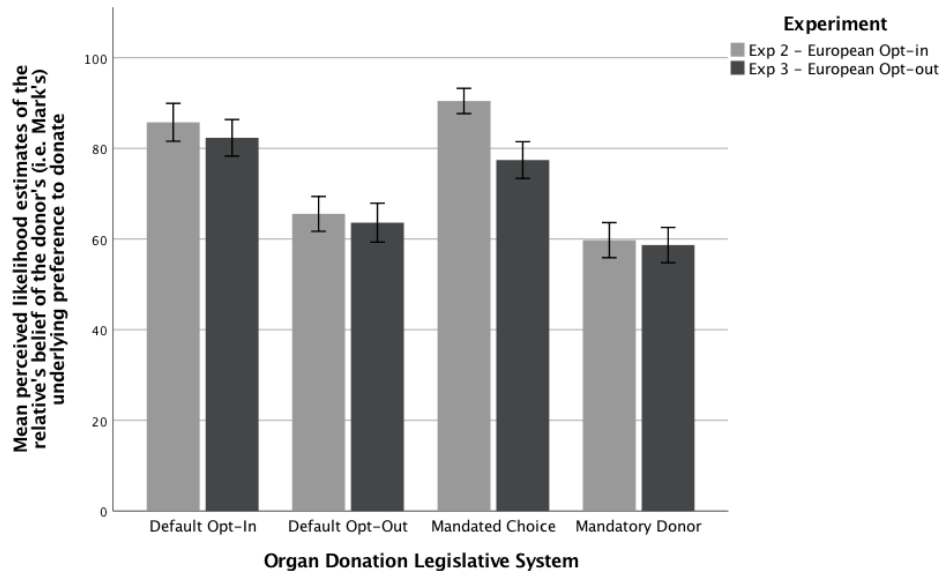


Figure 8. Mean perceived likelihood estimates of the relative's belief of the donor's (i.e. Mark's) belief of the donor's (i.e. Mark's) true preference to donate across the four experimental conditions for Experiment 2 and Experiment 3. Error bars are at 95% CI.

For Experiment 2, the perceived likelihood estimates of the relative's belief of the donor's true preference to donate decreased in the following order: Mandated choice = Default opt-in > Default opt-out > Mandatory donor. Planned pairwise comparisons confirmed this pattern: Perceived likelihood estimates of the relative's belief of the donor's true preference to donate were higher in the default opt-in system compared to the default opt-out system [*Experiment 2*: $t(199) = 7.06, p < .001, d = 0.99$]; and the mandatory donor system [*Experiment 2*: $t(197) = 9.06, p < .001, d = 1.28$]. Similarly, the default opt-out system had higher perceived likelihood estimates compared with the mandatory donor system [*Experiment 2*: $t(198) = 2.10, p = .037, d = 0.30$]. The mandated choice system had higher perceived likelihood estimates when compared with the default opt-out system [*Experiment 2*: $t(181) = 10.41, p < .001, d = 1.47$] and the mandatory donor system [*Experiment 2*: $t(178) = 12.78, p < .001, d = 1.81$]. There was

no significant difference between the default opt-in system and the mandated choice system [*Experiment 2*: $t(174) = 1.86, p = .064, d = 0.26$].

For Experiment 3, the perceived likelihood estimates of the relative's belief of the donor's true preference decreased in the following order: Mandated choice = Default opt-in > Default opt-out = Mandatory donor. Planned pairwise comparisons confirmed this pattern: Perceived likelihood estimates of the relative's belief of the donor's true preference to donate were higher in the default opt-in system compared to the default opt-out system [*Experiment 3*: $t(396) = 6.43, p < .001, d = 0.90$]; and the mandatory donor system [*Experiment 3*: $t(396) = 8.18, p < .001, d = 1.18$]. The mandated choice system has higher perceived likelihood estimates compared with the default opt-out system [*Experiment 3*: $t(396) = 4.75, p < .001, d = 0.66$] and the mandatory donor system [*Experiment 3*: $t(396) = 6.50, p < .001, d = 0.93$]. There was no significant difference for perceived likelihood estimates between the default opt-in system and mandated choice system [*Experiment 3*: $t(396) = 1.69, p = .092, d = 0.24$]; and between the default opt-out system and mandatory donor system [*Experiment 3*: $t(396) = 1.71, p = .088, d = 0.24$].

Perceived likelihood estimates of the relative's decision to consent

The findings broadly supported our second hypothesis such that perceived likelihood estimates of the relative's decision to consent to donate their family member's organs reflected sensitivity to the four different organ donation legislative systems (Figure 9). Overall, the evidence revealed that the organ donation legislative system had a significant effect on participants' perceived likelihood estimates of the relative's decision to consent to donate [*Experiment 1*: $F(3, 492) = 21.78, p < .001, \eta^2 = .12$; *Experiment 2*: $F(3, 400) = 22.12, p < .001, \eta^2 = .14$; *Experiment 3*: $F(3, 399) = 5.88, p = .001, \eta^2 = .08$].

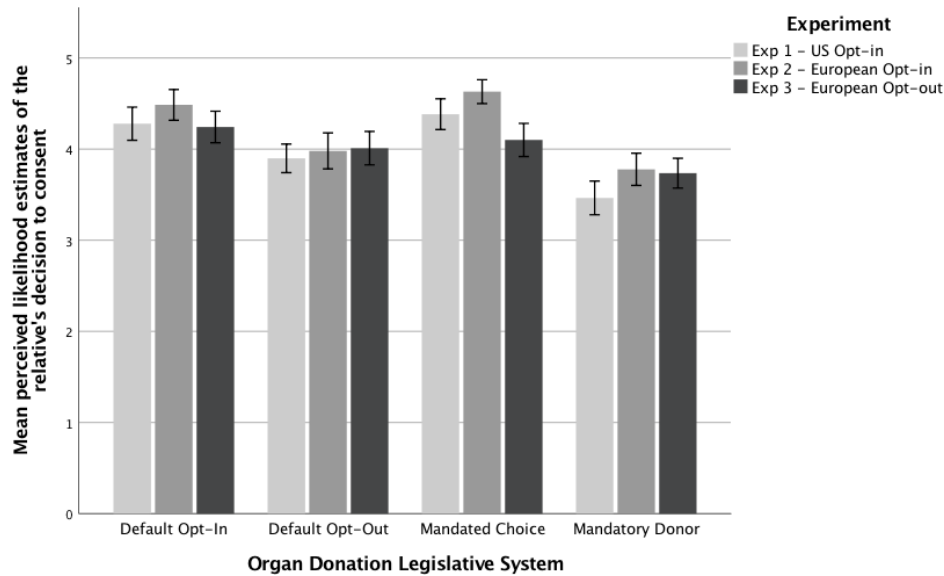


Figure 9. Perceived likelihood estimates of the relative's decision to consent across the four experimental conditions for Experiment 1, 2 and 3. Error bar at 95% CI.

In Experiment 1, the perceived likelihood estimates of the relative's decision to consent to donate their family member's organs decreased in the following order: Mandated choice = Default opt-in > Default opt-out > Mandatory donor. This is consistent with the pattern found for the perceived belief of the donor's true preference to donate. Planned pairwise comparisons confirmed this pattern: Participants' perceived likelihood estimates of the relative's decision to consent to donate were significantly higher under the default opt-in system compared to the default opt-out system [*Experiment 1*: $t(489) = 3.02, p = .003, d = 0.40$] and the mandatory donor system [*Experiment 1*: $t(489) = 6.57, p < .001, d = 0.79$]. Similarly, perceived likelihood estimates were higher under the mandated choice system compared to the default opt-out system [*Experiment 1*: $t(489) = 3.90, p < .001, d = 0.53$] and the mandatory donor system [*Experiment 1*: $t(489) = 7.50, p < .001, d = 0.92$]. The default opt-out system was also perceived as more likely to lead to family consent than the mandatory donor system [*Experiment 1*: $t(489) = 3.47, p = .001, d = 0.45$]. There was no significant difference in perceived likelihood estimates when comparing the mandated choice system and the default opt-in systems [*Experiment 1*: $t(489) = .84, p = .399, d = 0.11$].

For Experiment 2, the perceived likelihood estimates of the relative's decision to consent to donate their family member's organs decreased in the following order: Mandated choice = Default opt-in > Default opt-out = Mandatory donor. Planned pairwise comparisons confirmed

this pattern: Participants' perceived likelihood estimates of the relative's decision to consent to donate were significantly higher under the default opt-in system compared to the default opt-out system [*Experiment 2*: $t(195) = 3.86, p < .001, d = 0.55$] and the mandatory donor system [*Experiment 2*: $t(197) = 5.74, p < .001, d = 0.81$]. Similarly, perceived likelihood estimates were higher under the mandated choice system compared to the default opt-out system [*Experiment 2*: $t(174) = 5.44, p < .001, d = 0.77$], and the mandatory donor system [*Experiment 2*: $t(181) = 7.68, p < .001, d = 1.09$]. There was no significant difference in perceived likelihood estimates of the relative's decision to consent when comparing the mandated choice system and the default opt-in system [*Experiment 2*: $t(188) = 1.34, p = .181, d = 0.19$]; and between the default opt-out system and the mandatory donor system [*Experiment 2*: $t(196) = 1.52, p = .131, d = 0.21$].

In Experiment 3, the perceived likelihood estimates of the relative's decision to consent to donate their family member's organs decreased in the following order: Mandated choice = Default opt-in = Default opt-out > Mandatory donor. Planned pairwise comparisons confirmed this pattern: Participants' perceived likelihood estimates of the relative's decision to consent to donate were significantly higher under the default opt-in system compared to the mandatory donor system [*Experiment 3*: $t(396) = 4.06, p < .001, d = 0.59$]. Similarly, perceived likelihood estimates were higher under the mandated choice system compared to the mandatory donor system [*Experiment 3*: $t(396) = 2.93, p = .004, d = 0.41$]. The default opt-out system has a higher perceived likelihood estimates than the mandatory donor system [*Experiment 3*: $t(396) = 2.20, p = .028, d = 0.31$]. There was no significant difference in the perceived likelihood estimate of relative's decision to consent when comparing the mandated choice system and the default opt-in system [*Experiment 3*: $t(396) = 1.14, p = .257, d = 0.16$], and between the default opt-in system and the default opt-out system [*Experiment 3*: $t(396) = 1.85, p = .065, d = 0.26$]; as well as between the mandated choice system and the default opt-out system [*Experiment 3*: $t(396) = .72, p = .474, d = 0.10$].

The type of organ donation legislative systems was also significantly related to the binary decision of whether or not participants thought the relative would donate their deceased family member's organs: *Experiment 1* (the only experiment in which this question was included), $\chi(3) = 18.71, p < .001$. We present the percentage of participants that believe the relative would consent to donation under each organ donation legislative system: Default opt-in system

(92.6%), default opt-out system (94.9%), mandated choice system (96.1%), and mandatory donor system (82.4%). However, this measure was less sensitive, as compared to the other questions presented in the three studies.

Reason for Donating

In general, the organ donation legislative system in the participants' country (default opt-in or default opt-out) was significantly related to the reason they gave as to why the relative would consent to donate the deceased family member's organs (this question was not asked for *Experiment 1*, but was asked for *Experiment 2* and *Experiment 3*), $\chi(9) = 23.90, p = .004$; but not significantly related to the reasons for the relatives veto decision, $\chi(13) = 17.55, p = .176$. Table 2 and Table 3 show the frequency of each of the reasons chosen for consenting or vetoing decisions under each organ donation legislative system.

Table 2.

The combined general reasons in frequencies for opt-in and opt-out countries that John will consent to donate Mark's organs in each organ donation legislative system, where 'x' means this option was not displayed for that system.

Reason for likely/moderately likely/highly likely to donate	Opt-In	Opt-Out	Mandated Choice	Mandatory Donor
It is important to respect the deceased's wish	50	22	64	24
This is a gift of life	8	23	15	33
The act of good citizenship	7	18	7	39
There is evidence to suggest that Mark wanted to donate his organs	32	20	40	13
He actively made a choice to opt-in and donate his organs	95	10	68	16
Mark didn't opt-out during his lifetime means he wanted to donate	x	85	x	x
Mark couldn't opt-out of becoming a donor anyway	x	x	x	56
None of the above	0	7	1	9

Table 3.

The combined general reasons in frequencies for opt-in and opt-out countries that John will refuse to donate Mark's organs in each organ donation legislative system, where 'x' means this option was not displayed for that system.

Reason for moderately unlikely/highly unlikely to donate	Opt-In	Opt-Out	Mandated Choice	Mandatory Donor
It is a highly traumatic time for relatives and it is just not something they can think about	1	5	3	1
Lack of understanding of the organ donation process	0	1	0	2
Denial and rejection of brain-death criteria	0	1	0	1
The hope for a miracle	0	2	0	0
Fear about organ trade and unknown organ destination	1	0	0	0
Religious belief	2	0	0	0
Insecurity about the brain-death diagnosis	1	0	0	0
Unsure about Mark's wish to donate	0	4	1	2
Belief in body integrity	1	0	0	1
Fear about objection by other family members	0	0	0	0
There is a lack of evidence to indicate that Mark wanted to donate his organs	2	1	1	2
Because Mark didn't opt-out does not mean he wanted to donate his organs	x	1	x	x
Mark was forced to donate his organs	x	x	x	2
Mark was forced to make a choice when he wasn't ready	x	x	0	x
None of the above	0	0	0	0

Perceived diagnosticity of people's intention to donate

A one-way ANOVA was conducted to analyse perceived diagnosticity of evidence inferred from the likelihood ratio $\frac{P(\text{registration}|\text{want to donate})}{P(\text{registration}|\text{DON'T want to donate})}$.

The findings broadly supported the hypothesis that perceived diagnosticity of people's intention to donate, as assessed by the revealed likelihood ratios, again, reflected sensitivities to the four different organ donation legislative systems (Figure 10). A one-way ANOVA revealed a significant effect of organ donation legislative system on perceived diagnosticity of people's intention to donate [*Experiment 1*: $F(3, 415) = 9.53, p < .001, \eta^2 = .065$; *Experiment 2*: $F(3, 304) = 14.82, p < .001, \eta^2 = .13$; *Experiment 3*: $F(3, 336) = 12.78, p < .001, \eta^2 = .103$].

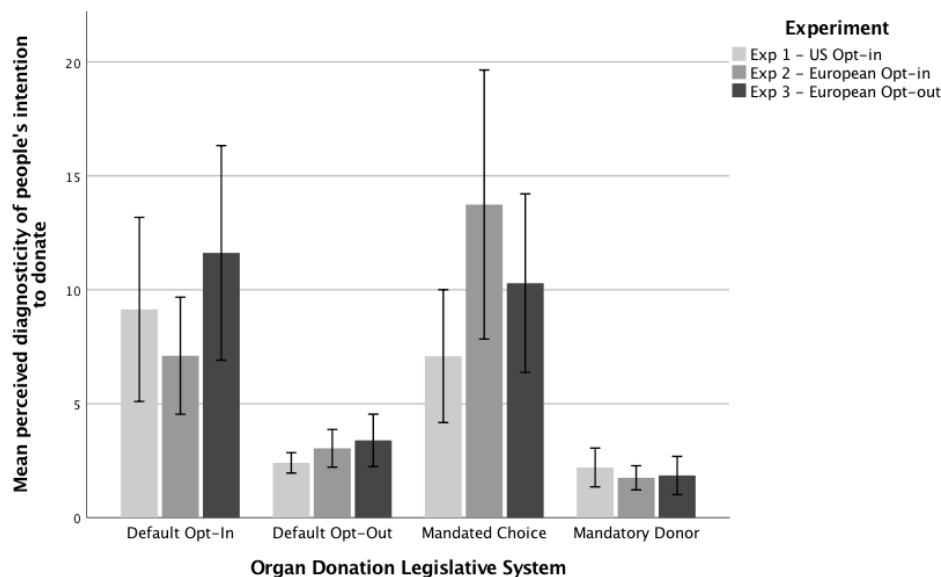


Figure 10. Mean perceived diagnosticity of people's intention to donate across the four experimental conditions for Experiment 1, 2 and 3. Error bar at 95% CI.

For Experiment 1, the perceived diagnosticity of people's intention to donate decreased in the following order: Mandated choice = Default opt-in > Default opt-out = Mandatory donor. Pairwise comparisons confirmed this pattern: The default opt-in system was perceived as more diagnostic of true preferences to donate than the default opt-out system [*Experiment 1*: $t(73) = 3.30, p = .001, d = 0.55$]; and the mandatory donor system [*Experiment 1*: $t(77) = 3.35, p = .001, d = 0.55$]. The mandated choice system was also perceived as more diagnostic than the default opt-out system [*Experiment 1*: $t(111) = 3.15, p = .002, d = 0.43$]; and the mandatory

donor system [*Experiment 1*: $t(124) = 3.19, p = .002, d = 0.43$]. There was no significant difference in perceived diagnosticity between the default opt-in system and the mandated choice system [*Experiment 1*: $t(140) = .82, p = .414, d = 0.13$]; and between the default opt-out system and the mandatory donor system [*Experiment 1*: $t(184) = .42, p = .679, d = 0.05$].

For Experiment 2, the perceived diagnosticity of people's intention to donate decreased in the following order: Mandated choice > Default opt-in > Default opt-out > Mandatory donor. Pairwise comparisons confirmed this pattern: The default opt-in system was perceived as more diagnostic of true preferences to donate than the default opt-out system [*Study 2*: $t(56) = 3.03, p = .004, d = 0.59$]; and the mandatory donor system [*Experiment 2*: $t(50) = 4.11, p < .001, d = 0.83$]. The mandated choice system was also perceived as more diagnostic than the default opt-in system [*Experiment 2*: $t(88) = 2.06, p = .042, d = 0.36$], the default opt-out system [*Experiment 2*: $t(69) = 3.58, p = .001, d = 0.62$]; and the mandatory donor system [*Experiment 2*: $t(67) = 4.04, p < .001, d = 0.70$]. Also, the default opt-out system was perceived as more diagnostic than the mandatory donor system [*Experiment 2*: $t(164) = 2.62, p = .010, d = 0.38$].

For Experiment 3, the perceived diagnosticity of people's intention to donate decreased in the following order: Mandated choice = Default opt-in > Default opt-out > Mandatory donor. Pairwise comparisons confirmed this pattern: The default opt-in system was perceived as more diagnostic of true preferences to donate than the default opt-out system [*Experiment 3*: $t(68) = 3.39, p = .001, d = 0.60$]; and the mandatory donor system [*Experiment 3*: $t(333) = 4.91, p < .001, d = 0.73$]. The mandated choice system was also perceived as more diagnostic than the default opt-out system [*Experiment 3*: $t(93) = 3.36, p = .001, d = 0.53$]; and the mandatory donor system [*Experiment 3*: $t(86) = 4.19, p < .001, d = 0.64$]. The default opt-out system was perceived as more diagnostic than the mandatory donor system [*Experiment 3*: $t(175) = .88, p = .033, d = 0.31$]. There was no significant difference between the default opt-in system and the mandated choice system [*Experiment 3*: $t(128) = .43, p = .666, d = 0.07$].

The impact of demographics on response patterns in Experiment 1-3

The demographic questions presented to all participants at the end of each experiment required that they indicate: 1) whether they were on the ODR, 2) whether they knew anyone that was on the ODR, 3) whether they were blood donors, 4) whether they would consent to donating their loved one's organs even if their wishes were unknown, 5) whether organ donation was

deemed as forbidden in their religion, and 6) whether they agree with the idea of a default opt-out system. In general, the key findings hold when including the above six factors as covariates. More specifically, the organ donation legislative system had a significant effect on participants' perceived belief of the donor's true preference to donate [*Experiment 1*: $F(3, 483) = 80.79, p < .001, \eta^2 = .04$; *Experiment 2*: $F(3, 391) = 48.30, p < .001, \eta^2 = .02$; *Experiment 3*: $F(3, 390) = 26.54, p < .001, \eta^2 = .01$]; perceived signal of donor's intent to donate [*Experiment 2*: $F(3, 391) = 97.26, p < .001, \eta^2 = .09$; *Experiment 3*: $F(3, 390) = 63.10, p < .001, \eta^2 = .07$]; likelihood estimates of the relative's belief of the donor's true preference to donate [*Experiment 2*: $F(3, 391) = 66.57, p < .001, \eta^2 = .03$; *Experiment 3*: $F(3, 390) = 31.79, p < .001, \eta^2 = .02$]; perceived likelihood of the relative's decision to consent to donate their family member's organs [*Experiment 1*: $F(3, 483) = 25.26, p < .001, \eta^2 = .008$; *Experiment 2*: $F(3, 391) = 23.77, p < .001, \eta^2 = .01$; *Experiment 3*: $F(3, 390) = 6.14, p < .001, \eta^2 = .002$]; and perceived diagnosticity of people's intention to donate [*Experiment 1*: $F(3, 406) = 9.82, \eta^2 = .03, p < .001$; *Experiment 2*: $F(3, 391) = 48.30, p < .001, \eta^2 = .10$; *Experiment 3*: $F(3, 295) = 13.12, p < .001, \eta^2 = .08$].

Across the three experiments ($N = 1,323$), the majority of the samples were non-religious (52%), much of the remainder identified themselves as Christian (40%). For 95% of the participants, organ donation was not forbidden in their religion. The sample also consisted of about 50% blood donors, 66% who knew someone who is registered as an organ donor, and 56% who agree with the idea of an opt-out system. In the event of death, participants agreed that the deceased person's wishes should be respected (75%), with a small minority believing that the family should have a final say (16%), and the remaining participants indicating they do not know (7%). Across the three experiments, it appears that less than 50% of the participants have knowledge of the current organ donation legislation in their country (Table 4).

Table 4.

Participants' knowledge of the current organ donation legislation in their country: Experiment 1 (US, opt-in system), Experiment 2 (European country, opt-in system), and Experiment 3 (European country, opt-out system).

	What is the current legislation in your country?		
	(%)		
	Experiment 1	Experiment 2	Experiment 3
Opt-out/Presumed consent	9.2	7.7	18.5
Mandatory/compulsory	1.8	0.5	1.3
Need to carry a donor card	0.8	21.9	9.0
Opt-in	42.8	44.6	18.8
Family or close friend will decide	8.7	3.7	8.8
Do not know	36.7	21.4	43.8

Regression Analysis. Regression Analysis were conducted to determine the proportion of variation in the dependent variables (perceived belief of the donor's true preference to donate, perceived signal of intention to donate, perceived likelihood estimates of the relative's belief of the donor's true preference to donate, perceived likelihood estimates of the relative's decision to consent) explained by the independent variables (Experiment, legislative system, age, gender, whether organ donation is forbidden in their religion, whether they know anyone who is registered as a donor, whether they are a blood donor, whether they agree with an opt-out system, whether they would donate their loved one's organs, and who should decide for the donor).

For perceived belief of the donor's true preference to donate, a significant regression equation was found $F(12, 559) = 11.19, p < .001$, with an R^2 of .197 (Table 5). Participants' ratings of perceived belief of the donor's true preference to donate decreased by 18.74 for the default opt-out system, and increased by 8.90 when the deceased decides for donation.

Table 5.

Summary of Multiple Regression of mean perceived belief of the donor's true preference to donate with predictor variables experiment, legislative system, age, gender, forbidden organ donation, organ donor, blood donor, known registered donor, agree with opt-out, willingness to donate loved one's organ, who decides in the event of death.

Variables	B	SE_b	β
Intercept	51.618	10.577	
Gender	2.263	1.765	0.052
Age	0.145	0.077	0.077
Forbidden Organ Donation	7.472	5.079	0.057
Organ Donor	(2.273)	2.074	(0.052)
Blood Donor	(0.686)	1.707	(0.016)
Known Organ Donor	1.385	2.035	0.031
Agree with Opt-out System	3.103	1.952	0.067
Donate Loved One's Organ	3.386	2.215	0.064
Experiment 2	0.39	1.883	0.009
Mandated Choice System	(2.366)	2.046	(0.051)
Default Opt-out System	(18.738)	2.049	(0.409)***
Deceased Decides Donation	8.898	2.379	0.147***

Note. $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; B = unstandardized coefficient; SE_b = Standard error of the coefficient; β = standard coefficient.

For perceived signal of the donor's intention to donate, a significant regression equation was found $F(12, 362) = 14.88$, $p < .001$, with an R^2 of .328 (Table 6). Participants' ratings of perceived signal of intention increased by 6.8 for the European opt-in countries, and decreased by 2.48 for mandated choice system and 33.32 for default opt-out systems.

Table 6.

Summary of Multiple Regression of mean perceived signal of the donor's intention to donate with predictor variables experiment, legislative system, age, gender, forbidden organ donation, organ donor, blood donor, known registered donor, agree with opt-out, willingness to donate loved one's organ, who decides in the event of death.

Variables	B	SE_b	β
Intercept	55.443	15.345	
Gender	2.227	2.514	.039
Age	.060	.108	.025
Forbidden Organ Donation	9.775	7.369	.058
Organ Donor	(1.565)	2.940	(.027)
Blood Donor	1.201	2.516	.021
Known Organ Donor	(2.372)	2.941	(.041)
Agree with Opt-out System	3.382	2.796	.057
Donate Loved One's Organ	.083	3.234	.001
Experiment 2	6.803	2.771	.118*
Mandated Choice System	(2.484)	3.032	(.041)*
Default Opt-out System	(33.322)	2.990	(.560)***
Deceased Decides Donation	5.627	3.476	.072

Note. $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; B = unstandardized coefficient; SE_b = Standard error of the coefficient; β = standard coefficient.

For perceived likelihood estimates of the relative's belief of the donor's true preference to donate, a significant regression equation was found $F(12, 547) = 14.86, p < .001$, with an R^2 of .246 (Table 7). Participants' ratings of perceived likelihood estimates increased with age by .15, by 16.16 when organ donation is forbidden in their religion, by 4.82 when the deceased decides for donation, decreased by 19.83 for a default opt-out system.

Table 7.

Summary of Multiple Regression of mean perceived likelihood estimates of the relative's belief of the donor's true preference to donate with predictor variables experiment, legislative system, age, gender, forbidden organ donation, organ donor, blood donor, known registered donor, agree with opt-out, willingness to donate loved one's organ, who decides in the event of death.

Variables	B	SE_B	β
Intercept	34.934	10.23	
Gender	3.005	1.707	0.069
Age	0.149	0.074	0.079
Forbidden Organ Donation	16.158	4.913	0.124**
Organ Donor	(0.587)	2.006	(0.013)
Blood Donor	(0.362)	1.651	(0.008)
Known Organ Donor	(0.909)	1.968	(0.02)
Agree with Opt-out System	1.802	1.888	0.039
Donate Loved One's Organ	3.257	2.143	0.061
Experiment 2	3.277	1.821	0.076
Mandated Choice System	0.323	1.979	0.007
Default Opt-out System	(19.832)	1.982	(0.434)***
Deceased Decides Donation	4.824	2.301	0.08*

Note. $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; B = unstandardized coefficient; SE_B = Standard error of the coefficient; β = standard coefficient.

For perceived likelihood estimates of the relative's decision to consent, a significant regression equation was found $F(12, 392) = 6.64$, $p < .001$, with an R^2 of .127 (Table 8). Participants' ratings of perceived likelihood estimates increased with age by .01, by .76 when organ donation is forbidden by religion, by .33 when the deceased decides for donation, and decreased by .38 for a default opt-out system.

Table 8.

Summary of Multiple Regression of mean perceived likelihood estimates of the relative's decision to consent with predictor variables experiment, legislative system, age, gender, forbidden organ donation, organ donor, blood donor, known registered donor, agree with opt-out, willingness to donate loved one's organ, who decides in the event of death.

Variables	B	SE_e	β
Intercept	1.958	0.456	
Gender	0.088	0.076	0.049
Age	0.011	0.003	0.139*
Forbidden Organ Donation	0.757	0.219	0.141*
Organ Donor	(0.069)	0.089	(0.038)
Blood Donor	(0.051)	0.074	(0.029)
Known Organ Donor	0.022	0.088	0.012
Agree with Opt-out System	0.037	0.084	0.019
Donate Loved One's Organ	0.189	0.095	0.086
Experiment 2	0.093	0.081	0.052
Mandated Choice System	0.004	0.088	0.002
Default Opt-out System	(0.379)	0.088	(0.200)***
Deceased Decides Donation	0.325	0.102	0.130**

Note. $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; B = unstandardized coefficient; SE_e = Standard error of the coefficient; β = standard coefficient.

5.4. Discussion

The aim of this study was to examine the welfare argument, under which nudge undermines autonomy. To do so, the current study examined third-party judgements as a way of gaining insights into inferred true preference from nudged choices (under default opt-in or default opt-out systems) in the context of organ donation. The motivation for doing this was to simulate the kind of information families are faced with when making decisions as to whether to donate their deceased relative's organs, as a way to consider a possible reason for high family refusal rates, namely the strength of signals of true preference to donate that are being perceived from choices under different organ donation legislative systems.

In the main, the findings across three studies supported Hypothesis 1 which is that people are sensitive to the framing of organ donation legislative systems under which a decision to donate is made. More specifically, when an individual has been ‘defaulted’ into donation under a default opt-out or mandatory donor system, participants perceived the donor’s (i.e. Mark’s) true preference to donate as weaker than when that choice was made actively under a mandated choice or default opt-in system (Experiment 1-3). This pattern was consistent with the perceived signal of the donor’s (i.e. Mark) intent to donate (Experiment 2 and 3 only). Subsequently, the findings supported Hypothesis 2, that families making decisions on the basis of their beliefs about the deceased’s true preference to donate will be perceived by participants as more likely to agree to donate. Indeed, the same pattern was revealed in participants’ perceived likelihood estimates of the relative’s belief of the donor’s (i.e. Mark) true preference to donate (Experiment 2 and 3 only), and perceived likelihood estimates of the relative’s decision to consent to donate a family member’s organs (Experiment 1 and 2, except for Experiment 3 where no difference was found between default opt-in and default opt-out). That is, these perceived likelihood estimates were judged as weaker under a default opt-out or mandatory donor system compared to a default opt-in or mandated choice system. A multiple regression of the respective dependent variables also revealed that ratings decreased consistently when a default opt-out system is introduced. Lastly, consistent with the patterns found in Hypothesis 1 and Hypothesis 2, the perceived diagnosticity of this evidence (i.e. Mark’s registration on the ODR across the different organ donation legislative systems) for inferring donor’s true preference to donate was stronger when he was registered under the default opt-in or mandated choice system compared to default opt-out or mandatory donor systems (Experiment 1-3). From a welfare perspective, an autonomous decision is more likely to reflect one’s true preference.

Across the three studies, participants gave different reasons for the relative (i.e. John) consenting or vetoing donation under each organ donation legislative system. The general reason in the default opt-in and mandated choice systems were that “he actively made a choice to opt-in and donate his organs” and “it is important to respect the deceased’s wish”. The reason for participants consenting to donate in the default opt-out and mandatory donor system was “Mark didn’t opt-out during his lifetime means he wants to donate” and “Mark couldn’t opt-out of becoming a donor anyway” respectively. This shows that, among the reasons for consenting to donation, participants were sensitive to the different systems under which the

donor (i.e. Mark's) was registered to donate. Chapter 7 will discuss the implication of these findings in further details, with particular focus on the debate between active choosing and defaults on welfare grounds.

CHAPTER 6: Public Attitude Towards Nudge

The focus of this chapter is to examine the extent to which nudge undermines autonomy if people cannot discern the intervention and how it is used to change their behaviour – the transparency argument. I explained in Chapter 3.2 that the mechanism by which nudge works, by triggering behavioural change without the chooser's awareness, poses profound questions on the ethics of nudging. In particular, nudges that lack transparency are regarded as manipulative and manipulation is objectionable because it violates people's autonomy by subverting the decision-making process. To ensure transparency, it has been stated that a watchful individual should possess the autonomous power to be able to discern the intention of the change in the choice architecture and to opt out of any arrangements if he/she wanted to. Given that previous survey findings have shown that people generally prefer transparent nudges over non-transparent nudges, there is speculation that public attitudes are influenced by the transparency of interventions. However, to date, there has not been explicit examination of the transparency of interventions to see how it influences people's judgments on a variety of dimension which are tested in this study. This motivated the current study to investigate the extent to which public attitudes towards four factors, namely, ease of identification, willingness to change behaviour, perceived effectiveness and acceptability, are dependent on transparency of an intervention. In addition, this study also examines the extent to which these four factors are dependent on the agent that propose the intervention and the presentation of rationale for these interventions across five contexts. This study will be conducted using UK and US participants. These two countries have been found to have different views on the use of restrictive behaviour change interventions (Branson et al., 2012). Because nudges have been extensively applied in both of these countries, the current study will explore the extent to which public opinion differ between these two countries across the five behaviour contexts.

The specific hypotheses are investigated here:

H1 (transparency). Across the five contexts, the public will find it easier to identify the actual behavioural intervention and how their behaviour will be changed by the psychological method in the transparent condition and in turn, more accurately classify these as transparent interventions.

H2 (transparency). Across the five contexts, the public will be more willing to change their behaviour when presented with transparent nudges compared to non-transparent nudges.

H3 (transparency). Across the five contexts, the public will generally perceive an intervention as more effective in the transparent condition than the non-transparent condition.

H4 (transparency). Across the five contexts, the public will generally perceive an intervention as more acceptable in the transparent condition than the non-transparent condition; and there will be a correlation between perceived effectiveness and acceptability.

H5 (transparency). Public opinion with regard to ease of identification, perceived effectiveness and acceptability of an intervention, and their willingness to change will be influenced by the presentation of rationale for these interventions and judgements will follow the direction of: positive > positive + negative > negative.

H6 (transparency). Public opinion with regard to ease of identification, perceived effectiveness and acceptability of an intervention, and their willingness to change will be influenced by the behaviour agent that proposed the intervention, such that judgments overall will be most favourable towards scientists proposing nudges as compared to advertisers and policy makers.

6.1. Participants and Design

In each experiment, there were two samples (Table 9), Experiment 1 (US $N = 265$, UK $N = 237$), Experiment 2 (US $N = 306$, UK $N = 305$), and Experiment 3 (US $N = 301$, UK $N = 301$). All Experiments were presented via Qualtrics and launched via Prolific Academic a crowd sourcing system for participant recruitment worldwide, and all were financially compensated for their time (£1.21). The three experiments differed in terms of the rationale presented for the interventions: Positive argument that highlight why the intervention works (Experiment 1) negative arguments that highlight why the intervention does not work (Experiment 3), and presentation of both positive and negative arguments (Experiment 2). For full details of the questionnaires and specific instructions, please see Appendix D.

The design of the experiment was a 3 (Argument: Positive, Positive & Negative, Negative) x 2 (Sample: US, UK) x 3 (Behavioural Agent: Scientists, Government, Advertisers) x 2 (Transparency: Transparent, Non-transparent) as between-subject variables, x 5 (Contexts:

Smoking, Diet, Alcohol, Exercise, Banking) as the within-subject variable. They were given six probative questions regarding behavioural interventions, and five demographic questions⁶.

Table 9.

Participants profile from Experiment 1, 2 and 3 combined

Sample	US	UK
Total participants	Total combining Experiment 1, 2 and 3 $N = 872$ (all US residents or nationals, first language English)	Total combining Experiment 1, 2 and 3 $N = 843$ (all UK residents or nationals, first language English)
Females	471 (54%)	413 (49%)
Age	Mean 35 ($SD = 12.24$) ranging from 18-74	Mean 32.36 ($SD = 11.32$) ranging from 18-71
Educational background	Mixed, 56.7% qualified with a degree (at bachelor degree and postgraduate level).	Mixed, 57.1% qualified with a degree (at bachelor degree and postgraduate level).
Political affiliation	51.6% identifying as left, 8.6% as centre, 16.7% as right, and 23.1% as other	47% identifying as left, 16.1% as centre, 17% as right, and 19.9 % as other
Religion	41.6% reported that they did not have one, 19.5% were not sure, 38.9% reported that they were religious	38.9% reported that they did not have one, 23.8% were not sure, 37.3% reported that they were religious

⁶ With regards to the demographic questions, participants were asked to type their age in a text box [continuous measure], and provide details of their gender (Female [score 1], Male [score 2]), along with their political affiliation (Left [Score 1], Centre [Score 2], Right [Score 3], Other [Score 4]), their educational background (High school [Score 1], diploma/foundation [Score 2], Bachelors degree [Score 3], Master's degree [Score 4], PhD [Score 5], Other [Score 6]), and religious orientation (Religious [Score 1], Not sure [Score 2], Not religious [Score 3]). These formed the basic question regarding demographics and social/cultural characteristics.

Table 10.

Description of Transparent and Non-Transparent Behavioural Interventions.

Context	Transparent	Non-transparent
Smoking	Design cigarette packaging so that it incorporates graphic pictures of damaged lungs and warnings such as “Smoking seriously harms you and others around you”, “Smoking harms your unborn baby”.	Increasing the length of the filter by 10mm and at the same time reduce the length of the cigarette to 60mm.
Food	Design packaging on food so that the front label includes nutritional information, by using a simple traffic light system (red, amber, green) to indicate how much saturated fat, salt and sugar, and calories are in food products.	Design the size of plates so that the quantity of food on them is adjusted. Large plates and bowls can make servings of food appear smaller, whereas smaller plates can lead people to misjudge that very same quantity of food as being significantly larger.
Alcohol	Design signage in pubs and restaurants so that they include messages such as the following: “men and women are advised not to regularly drink more than 14 units a week” and “spread your drinking over three days or more if you drink as much as 14 units a week”.	Design the glassware used in pubs and restaurants in such a way so that straight glasses are used, because relative to curvy glasses, it is easier to judge and pace the amount of alcohol consumed.
Exercise	Design stairwells with ‘point-of-choice’ signage that displays messages about the health advantages of taking the stairs, such as “Stair climbing burns more calories per minute than tennis”, “7 minutes of stair climbing per day protects your heart”, etc.	Design stairwells by hanging artworks. Pictures are changed periodically to keep stair users to prolong effectiveness.
Finance	Design investment schemes in such a way so that customers can evaluate the associated riskiness of each product based on a traffic light system; red indicates highly risky, green indicates low risk.	Design investment schemes with an automatic enrolment system so Bank/Building Society will decide on an individual’s behalf exactly how the money will be allocated to investment schemes. Although if the individual didn’t want it, they could opt-out of the scheme, this would involve filling in relevant paperwork.

6.2. General Materials and Procedure

Participants were provided with a brief introduction to the use psychological methods used to bring about behaviour changes, and were told that these methods are designed to help guide people to make better decisions for their own health and wellbeing. First, they were randomly assigned to either the transparent or non-transparent condition, then they read a passage about the expert that proposed the intervention: “*The X in this country is using psychological research to help develop a set of simple methods that adjust the way information is presented, so that it can help people to make better decisions. The reason for using psychological methods is to help improve people’s behaviour, because in many day-to-day contexts people may not make a decision that is best for their own health, wellbeing, and their happiness.*”, where *X* represents either the Government, Top Advertising Company or Top Researchers in laboratories. They then read about the contexts (i.e. Smoking, Food, Alcohol, Exercise, and Banking) for which the psychological method has been implemented (Table 10). Depending on which experiment participants took part in (i.e., 1, 2, or 3), the rationale for why the interventions work or not work were presented. These were either a positive argument for why nudges work (e.g., by highlighting the negative physical and moral issues concerning smoking, the negative experiences will become more obviously associated with smoking and this will encourage smokers to reduce or even stop smoking), a negative argument for why they do not work (e.g., by highlighting the negative physical and moral issues concerning smoking, smokers will feel more defensive of their smoking habits, and as a result, smokers will end up smoking more, meaning that the method will lead to increases in smoking), and both positive and negative arguments. These arguments are specific to the psychological methods used in each context. Subsequently, they are presented with the definition of two types of psychological methods: Transparent and non-transparent. A transparent nudge method worked in such a way that anyone can easily identify the psychological method used to change their behaviour, as well as easily identify how their behaviour is changed by it. A non-transparent nudge worked in such a way that no one can easily identify the psychological method used to change their behaviour, and that no one can easily identify how their behaviour is changed by it. Based on these definitions, participants were required to answer the following six questions for each of the five contexts:

1. To what extent is it easy to for you to identify the use of psychological methods to change your behaviour? [I cannot identify the psychological method used to change my behaviour¹ to I can easily identify the psychological¹⁰⁰]

2. To what extent is it easy for you to identify HOW your behaviour is changed by the psychological method? [I cannot identify how my behaviour is changed by the psychological method¹ to I can easily identify how many behaviour is changed by the psychological method¹⁰⁰]
3. Is the psychological method described above transparent or non-transparent? [Binary response: transparent/non-transparent]
4. To what extent do you want to change your behaviour through the psychological method in this particular situation? [Not at all¹ to Very Much⁹]
5. To extent do you think the psychological method described above would positively change YOUR behaviour? [Much less likely¹to Much more likely⁹]
6. To what extent do you think it is acceptable to use the psychological method described above to change your behaviour? [Unacceptable¹ toAcceptable⁹]

Responses to Question 1 and 2 were combined to obtain an average transparency rating of ease of identification of a nudge intervention. Question 3 elicited accuracy of identifying the intervention as either transparent or non-transparent depending on the conditions participants were assigned to. Questions 4 elicited people's willingness to change their behaviour in response to the implementation of a nudge intervention. Question 5 elicited perceived effectiveness of a nudge intervention. Question 6 elicited public acceptability of the nudges.

6.3.Results

Accuracy of Identification. The findings tested Hypothesis 1, in that the public will be more accurate at identifying transparent nudges compared to non-transparent nudges. Across the five contexts, Table 11 shows that participants are generally accurate at identifying transparent interventions as transparent but are less accurate at identifying non-transparent interventions (Smoking: $\chi^2(1) = 701.86, p < .001$; Food: $\chi^2(1) = 458.57, p < .001$; Alcohol: $\chi^2(1) = 185.16, p < .001$; Exercise: $\chi^2(1) = 121.59, p < .001$; Finance: $\chi^2(1) = 49.73, p < .001$).

Table 11.

Contingency table showing people's accuracy in identifying the transparency of an intervention across the five behavioural contexts.

Transparency of Intervention	Smoking		Food		Alcohol		Exercise		Finance	
	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Transparent (%)	94.8	5.2	83.7	16.3	82.3	17.7	84.6	15.4	70.7	29.3
Non- Transparent (%)	33.4	66.6	32.7	67.3	51.3	48.7	60.9	39.1	54.3	45.6

Note. ✓ = Participants have correctly identified the intervention under the transparent and non-transparent condition respectively; ✗ = Participants have incorrectly identified the intervention under the transparent and non-transparent condition respectively.

To examine the impact of the manipulations on judgments regarding the ease of identifying the actual behavioural intervention designed to change behaviour, a mixed ANOVA was conducted on a 3(Argument: Positive, Positive & Negative, Negative) x 2 (Sample: US, UK) x 3 (Behavioural Agent: Scientists, Government, Advertisers) x 2 (Transparency: Transparent, Non-transparent) as between-subject variables, x 5 (Contexts: Smoking, Diet, Alcohol, Exercise, Banking) as the within-subject variable. For the purposes of a clean presentation, only effect sizes greater or equal to 0.01 are reported. Cohen (1988) provided benchmarks to define partial η_p^2 (R-squared in a multiple regression) of small (0.01), medium (0.09) and large (0.25) effects, any effects lower than 0.01 are not reported.

Ease of Identification. The findings supported Hypothesis 1 such that the ease of identification is higher for the transparent condition ($M = 72.98$, $SD = 21.10$) than the non-transparent condition ($M = 60.36$, $SD = 21.14$), $F(1, 1679) = 299.32$, $p < .001$, $\eta_p^2 = .15$, 99% $CI [10.74, 14.50]$.

The findings supported Hypothesis 5 such that the presentation of rationale for the interventions influenced people's judgement of ease of identification, $F(2, 1679) = 6.72$, $p = .001$, $\eta_p^2 = .01$. Pairwise comparison was applied to all follow-up analyses. When a positive rationale was presented ($M = 68.10$, $SD = 27.54$), participants perceived an intervention as easier to identify compared to a negative rationale ($M = 64.87$, $SD = 24.91$), $t(1104) = 3.55$, $p = .001$, 99% $CI [.56, 5.90]$; a combination of a positive & negative rationale ($M = 67.03$, $SD = 24.91$) also lead to higher ease of identification than a negative rationale ($M = 64.87$, $SD = 25.12$), $t(1213) = 2.50$, $p = .037$, 99% $CI [-.38, 4.70]$. There were no significant differences in

ease of identification between a positive rationale and a combined positive & negative rationale. Overall, independent of transparency, judgement of ease of identification for rationale of intervention presented decrease in the following direction: positive = positive & negative > negative. Hypothesis 6 was not supported given that participants' judgement of ease of identification was not influence by the agent that proposed the intervention.

There was a main effect of context, $F(4, 1676) = 135.84, p < .001, \eta_p^2 = .08$, in which all paired comparisons between all five context revealed significant differences ($p < .001$) with perceived transparency decreasing in the following direction from most to least ease of identification: Diet > Smoking > Exercise > Alcohol > Banking (Figure 11). There was an interaction between transparency * context, $F(4, 1679) = 63.92, p < .001, \eta_p^2 = .04$ (Figure 12). Pairwise comparison across all five contexts revealed significant difference between the transparent and non-transparent condition for the smoking context, $t(1713) = 10.03, p < .001, d = .48$; the food context, $t(1713) = 3.81, p < .001, d = .18$; the alcohol context, $t(1713) = 11.29, p < .001, d = .55$; the exercise context, $t(1713) = 20.66, p < .001, d = 1.00$; and the finance context, $t(1713) = 10.03, p < .001, d = .48$.

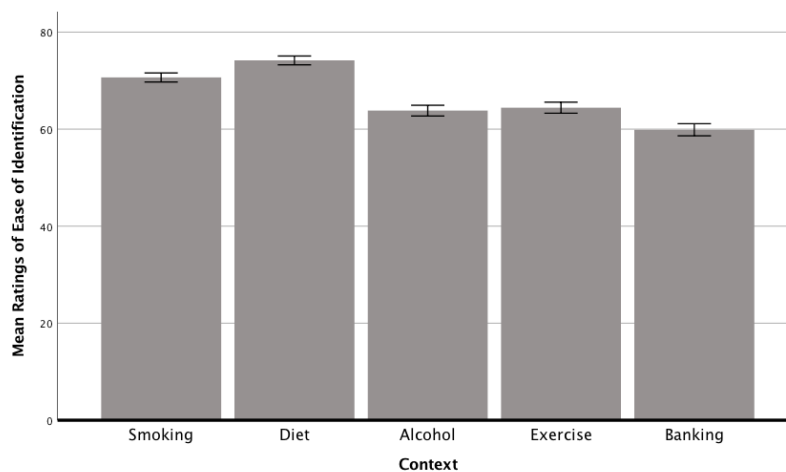


Figure 11. The main effect of context on mean ratings of ease of identification. Error bar at 99% CI.

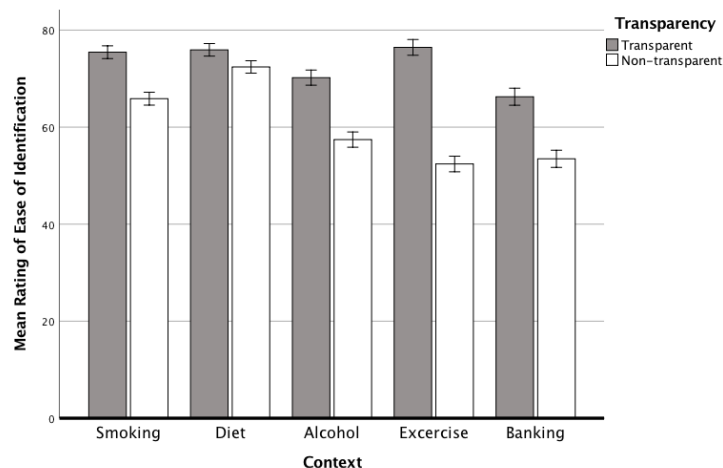


Figure 12. The interaction between context and transparency of an intervention on mean ratings of ease of identification. Error bar at 99% CI.

Willingness to Change Behaviour. The findings supported Hypothesis 2 such that participants were more willing to change their behaviour under transparent interventions ($M = 4.93$, $SD = 2.21$) compared to non-transparent interventions ($M = 4.27$, $SD = 2.21$), $F(1, 1679) = 73.51$, $p < .001$, $\eta_p^2 = .04$, 99% $CI [.46, .85]$.

Also, the findings supported Hypothesis 5 such that the presentation of a rationale for the intervention influenced people's willingness to change their behaviour, $F(2, 1679) = 9.14$, $p < .001$, $\eta_p^2 = .0$. Pairwise comparison revealed a higher willingness to change their behaviour under a positive rationale ($M = 4.83$, $SD = 2.87$) compared to a combined positive & negative rationale ($M = 4.53$, $SD = 2.62$), $t(1113) = 3.19$, $p = .004$, 99% $CI [.02, .58]$; and a negative rationale ($M = 4.44$, $SD = 2.62$), $t(1104) = 4.14$, $p < .001$, 99% $CI [.11, .67]$. No significant differences were found when comparing combined positive & negative rationales with a negative rationale. Overall, independent of transparency, judgement of willingness to change behaviour for rationale of intervention presented decreased in the following direction: positive > positive & negative = negative. Hypothesis 6 was not supported as the agent proposing the intervention did not influence participants' judgements of willingness to change their behaviour.

Looking specifically at context, there was a main effect of context, $F(4, 1679) = 249.25$, $p < .001$, $\eta_p^2 = .13$, in which all paired comparisons between all five context revealed significant differences ($p < .01$) with willingness to change behaviour decrease in the following direction

from most to least willing to change: Diet > Exercise > Smoking > Alcohol > Banking (Figure 13). There was an interaction between context * transparency, $F(4, 1679) = 34.72, p < .001, \eta_p^2 = .02$ (Figure 14). Pairwise comparison across all five contexts revealed significant difference between the transparent and non-transparent condition for the smoking context, $t(1713) = 1.95, p = .05, d = .10$; the food context, $t(1713) = 4.47, p < .001, d = .22$; the alcohol context, $t(1713) = 1.16, p = .25, d = .06$; the exercise context, $t(1713) = 7.95, p < .001, d = .38$; and the finance context, $t(1713) = 13.78, p < .001, d = .67$.

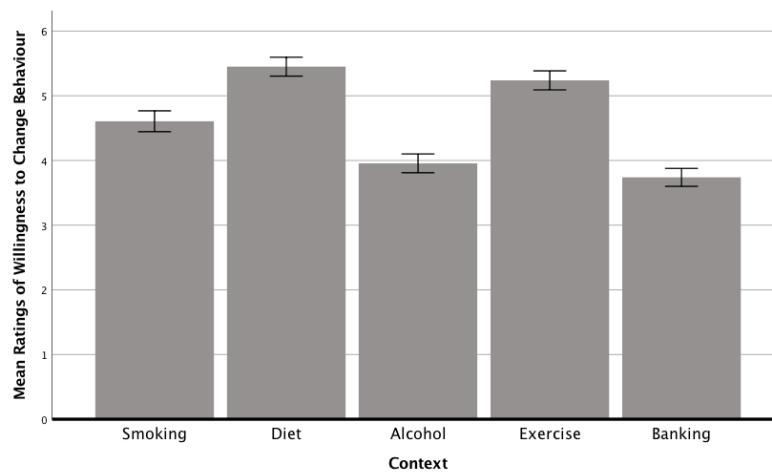


Figure 13. Main effect of context on mean ratings of willingness to change behaviour. Error bar at 99% CI.

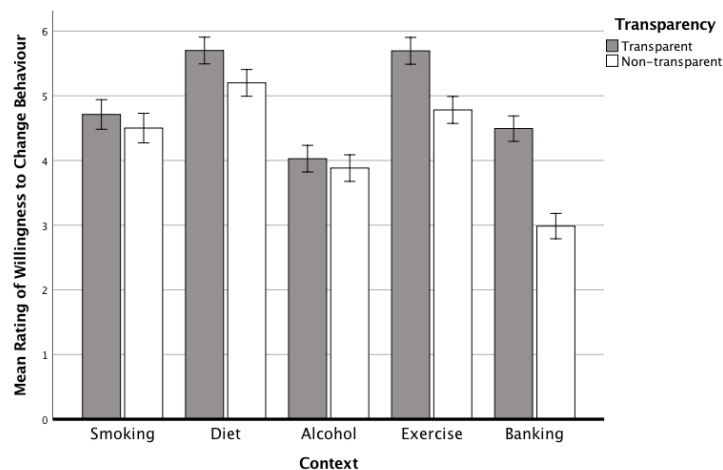


Figure 14. The interaction between context and transparency on the mean rating of willingness to change behaviour. Error bar at 99% CI.

Perceived Effectiveness. The findings supported Hypothesis 3 such that participants perceived an intervention as more effective when it was transparent ($M = 4.98, SD = 2.17$) than non-transparent ($M = 4.28, SD = 2.17$), $F(1, 1679) = 87.35, p < .001, \eta_p^2 = .05, 99\% CI [.51, .90]$.

The findings supported Hypothesis 5 such that perceived effectiveness of an intervention was judged differently when different rationales were presented for the intervention, $F(2, 1679) = 4.88, p = .008, \eta_p^2 = .01$. Pairwise comparison revealed that the perceived effectiveness of an intervention was higher when presented with a positive rationale ($M = 4.78, SD = 2.83$) compared with a negative rationale ($M = 4.49, SD = 2.58$), $t(1104) = 3.13, p = .006, 99\% CI [.02, .57]$. All other comparisons were not statistically significant. Overall, independent of transparency, judgement of perceived effectiveness for rationale of intervention presented decreased in the following direction: positive > negative. Hypothesis 6 was not supported as the agent that proposed the intervention did not influence participants' judgements of perceived effectiveness.

Overall, there was a main effect of context, $F(4, 1679) = 231.95, p < .001, \eta_p^2 = .12$. All paired comparisons between all five contexts revealed significant differences ($p < .05$), with perceived effectiveness of an intervention decreasing in the following direction from most to least perceived effectiveness: Diet > Exercise > Smoking > Alcohol > Banking (Figure 15). There was also a significant interaction between context * transparency, $F(4, 1679) = 36.63, p < .001, \eta_p^2 = .02$ (Figure 16). Pairwise comparison across all five contexts revealed significant difference between the transparent and non-transparent condition for the smoking context, $t(1713) = 4.73, p < .001, d = .23$; the food context, $t(1713) = 3.21, p = .001, d = .16$; the alcohol context, $t(1713) = 1.23, p = .22, d = .06$; the exercise context, $t(1713) = 9.08, p < .001, d = .44$; and the finance context, $t(1713) = 13.71, p < .001, d = .66$.

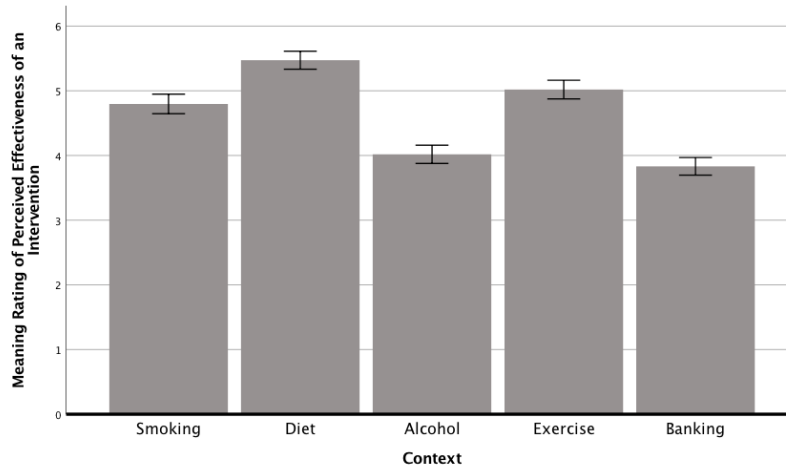


Figure 15. The main effect of context on mean ratings of perceived effectiveness of an intervention. Error bar at 99% CI.

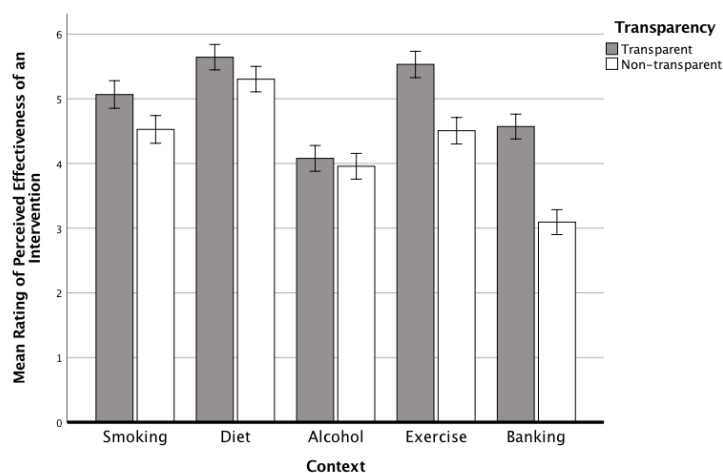


Figure 16. The interaction between context and transparency on mean ratings of perceived effectiveness of an intervention. Error bar at 99% CI.

Public Acceptability. The findings supported Hypothesis 3 such that participants found an intervention more acceptable when it was transparent ($M = 6.96$, $SD = 2.01$) than non-transparent ($M = 5.86$, $SD = 2.01$), $F(1, 1679) = 251.75$, $p < .001$, $\eta_p^2 = .13$, 99% CI [.92, 1.28]. In general, UK participants ($M = 6.53$, $SD = 2.05$) found an intervention more acceptable than US participants ($M = 6.30$, $SD = 1.97$), $F(1, 1679) = 11.75$, $p = .001$, $\eta_p^2 = .01$, 99% CI [.92, 1.28]. This was the only variable in which there were sample differences.

The findings supported Hypothesis 5 such that public acceptability was significantly different between different rationales of an intervention, $F(2, 1679) = 11.13, p < .001, \eta_p^2 = .01$. An intervention was perceived as more acceptable when presented under a positive rationale ($M = 6.65, SD = 2.62$) compared to a combined positive & negative rationale ($M = 6.33, SD = 2.38$), $t(1113) = 3.67, 99\% CI [.06, .57]$; and compared with a negative rationale ($M = 6.26, SD = 2.38$), $t(1104) = 4.51, 99\% CI [.13, .64]$. Overall, independent of transparency, judgement of perceived effectiveness for rationale of intervention presented decrease in the following direction: positive > positive & negative = negative. Hypothesis 6 was not supported as the agent that proposed the intervention did not influence people's judgments of acceptability.

Overall, there was a main effect of context, $F(4, 1679) = 612.50, p < .001, \eta_p^2 = .27$. Pairwise comparison between all contexts revealed significant differences ($p < .001$) except between smoking and alcohol. Public acceptability of an intervention from most to least acceptable is as follows: Exercise > Diet > Smoking = Alcohol > Banking (Figure 17). There was also an interaction between context * transparency, $F(4, 1679) = 175.13, p < .001, \eta_p^2 = .09$ (Figure 18). Pairwise comparison across all five contexts revealed significant difference between the transparent and non-transparent condition for the smoking context, $t(1713) = 6.75, p < .001, d = .32$; the food context, $t(1713) = 9.26, p < .001, d = .45$; the alcohol context, $t(1713) = 7.92, p < .001, d = .38$; the exercise context, $t(1713) = 2.25, p = .025, d = .11$; and the finance context, $t(1713) = 27.03, p < .001, d = 1.30$.

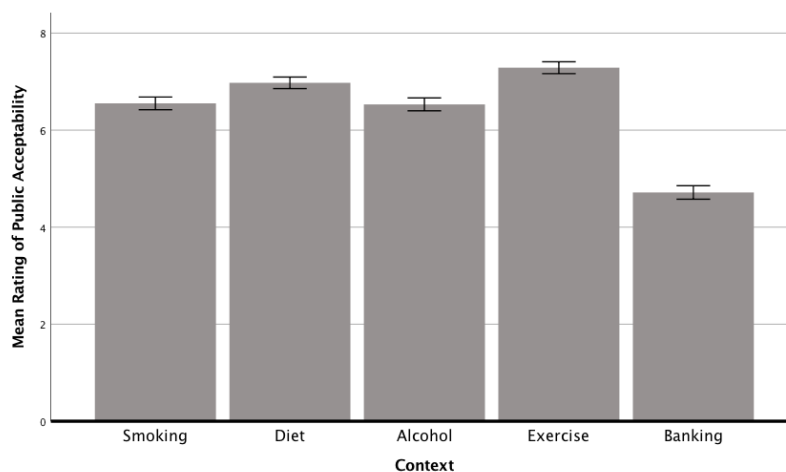


Figure 17. The main effect of context on mean rating of public acceptability. Error bar at 99% CI.

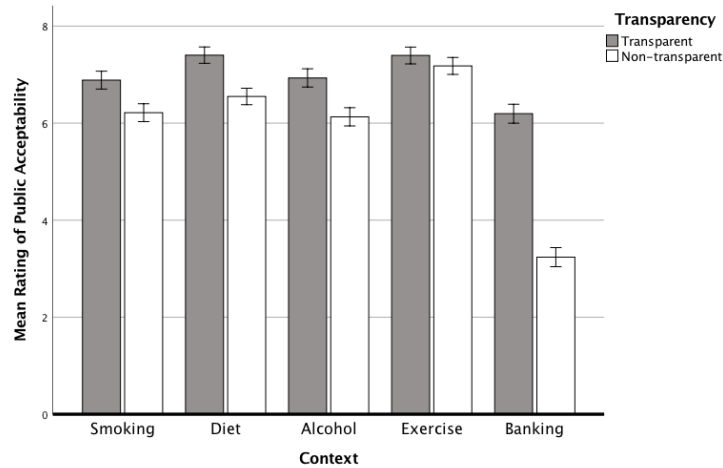


Figure 18. The interaction between context and transparency on mean rating of public acceptability. Error bar at 99% CI.

Regression Analysis. Regression analyses were conducted to determine the proportion of variation in the dependent variables (perceived transparency, willingness to change behaviour, perceived effectiveness and public acceptability) explained by the independent variables (transparency, behaviour agent, and rationale of experiment) and covariates (age, gender, political affiliation, religiosity, education level).

For ease of identification, a significant regression equation was found $F(14, 1691) = 22.69, p < .001$, with an R^2 of .158 (Table 12). Participants' ratings of *ease of identification* decreased by 12.45 when the intervention was non-transparent, decreased by 1.72 for US participants, and increased by 3.45 for positive rationales, and 2.25 for a combined positive & negative rationale.

Table 12.

Summary of Multiple Regression of mean ease of identification with predictor variables transparency, sample, behaviour agent, rationale of intervention, age, gender, political affiliation, religiosity, education level

Variables	B	SE_B	β
Intercept	72.641	1.723	
Sample	(1.72)	.743	(.052)*
Transparency	(12.45)	.733	(.38)***
Positive Rationale	3.453	.917	.096***
Postive+Negative Rationale	2.246	.872	.066*
Advertiser	(.768)	.897	(.022)
Government Official	(1.298)	.894	(.037)
Left Political Affiliation	(.663)	.954	(.02)
Right Political Affiliation	(.669)	1.2	(.015)
Centre Political Affiliation	(.771)	1.322	(.016)
Religious	(.564)	.832	(.017)
Not Religious	.202	.99	.005
Age	.028	.032	.02
Gender	(.831)	.742	(.025)
Education Level	.089	.45	.004

Note. $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; B = unstandardized coefficient; SE_B = Standard error of the coefficient; β = standard coefficient.

For ratings of *willingness to change behaviour*, a significant regression equation was found $F(14, 1691) = 10.62, p < .001$, with an R^2 of .081 (Table 13). Participants' rating of willingness to change behaviour decreased by .67 when the intervention is non-transparent, decreased by .185 when the intervention was proposed by Government officials and .20 by Advertisers, and increased by .39 in the presence of positive rationales, increased by .491 for religious participants, and decreased by .013 with increases in age.

Table 13.

Summary of Multiple Regression of mean willingness to change behaviour with predictor variables transparency, sample, behaviour agent, rationale of intervention, age, gender, political affiliation, religiosity, education level

Variables	B	SE_B	β
Intercept	5.129	.178	
Sample	.043	.077	0.013
Transparency	(.67)	.076	(.206)***
Positive Rationale	.387	.095	.108***
Postive+Negative Rationale	.089	.09	.026
Advertiser	(.2)	.093	(.058)*
Government Official	(.185)	.093	(.054)*
Left Political Affiliation	(.039)	.099	(.012)
Right Political Affiliation	.072	.124	.017
Centre Political Affiliation	.008	.137	.002
Religious	.491	.086	.147***
Not Religious	.149	.103	.038
Age	(.013)	.003	(.095)***
Gender	(.014)	.077	(.004)
Education Level	(.022)	.047	(.011)

Note. $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; B = unstandardized coefficient; SE_B = Standard error of the coefficient; β = standard coefficient

For ratings of *perceived effectiveness*, a significant regression equation was found $F(14, 1691) = 11.84, p < .001$, with an R^2 of .089 (Table 14). Participants' rating of perceived effectiveness decreased by .711 when the intervention is non-transparent, decreased by .244 when the intervention was proposed by Advertisers, and increased by .28 in the presence of a positive rationales, increased by .498 for religious participants, and decreased by .014 with increases in age.

Table 14.

Summary of Multiple Regression of mean perceived effectiveness with predictor variables transparency, sample, behaviour agent, rationale of intervention, age, gender, political affiliation, religiosity, education level

Variables	B	SE_B	β
Intercept	5.342	.174	
Sample	.058	.075	.018
Transparency	(.711)	.074	(.223)***
Positive Rationale	.28	.093	.08**
Postive+Negative Rationale	.111	.088	.033
Advertiser	(.244)	.091	(.072)**
Government Official	(.164)	.09	(.049)
Left Political Affiliation	(.077)	.096	(.024)
Right Political Affiliation	.052	.121	.012
Centre Political Affiliation	.018	.134	.004
Religious	.498	.084	.152***
Not Religious	.045	.1	.012
Age	(.014)	.003	(.107)***
Gender	(.085)	.075	(.027)
Education Level	(.046)	.045	(.023)

Note. $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; B = unstandardized coefficient; SE_B = Standard error of the coefficient; β = standard coefficient

For ratings of *public acceptability*, a significant regression equation was found $F(14, 1691) = 21.42, p < .001$, with an R^2 of .151 (Table 15). Participants' rating of public acceptability decreased by 1.10 when the intervention was non-transparent, decreased by .21 for US participants, decreased by .22 when the intervention was proposed by Government officials, and increased by .28 in the presence of a positive rationale, and decreased by .01 with increases in age.

Table 15.

Summary of Multiple Regression of mean public acceptability with predictor variables transparency, sample, behaviour agent, rationale of intervention.

Variables	B	SE _B	β
Intercept	7.18	.163	
Sample	(.208)	.071	(.067)**
Transparency	(1.099)	.07	(.355)***
Positive Rationale	.382	.087	.112***
Postive+Negative Rationale	.08	.083	.025
Advertiser	(.154)	.085	(.047)
Government Official	(.22)	.085	(.067)**
Left Political Affiliation	.049	.091	.016
Right Political Affiliation	.072	.114	.017
Centre Political Affiliation	.139	.125	.03
Religious	.042	.079	.013
Not Religious	.055	.094	.015
Age	(.006)	.003	(.048)*
Gender	(.065)	.07	(.021)
Education Level	.018	.043	.009

Note. $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; B = unstandardized coefficient; SE_B = Standard error of the coefficient; β = standard coefficient

Relationship with Public Acceptability

Given that previous studies have emphasized the importance of public acceptability, a significant regression equation was found for all three other factors. For *mean ease of identification*, a significant regression equation was found $F(1, 1713) = 427.44, p < .001$, with an R^2 of .20. Participants' rating of public acceptability increased by .042 as the intervention became easier to identify. For *mean willingness to change behaviour*, a significant regression equation was found $F(1, 1713) = 898.96, p < .001$, with an R^2 of .219. Participants' rating of public acceptability increased by .446 with an increase in mean willingness to change their behaviour. For *mean perceived effectiveness*, a significant regression equation was found $F(1, 1713) = 444.07, p < .001$, with an R^2 of .206. Participants' rating of public acceptability increased by .441 with an increase in mean perceived effectiveness of the intervention.

6.4. Discussion

The aim of the study was to empirically explore the extent to which nudge undermines perceived autonomy if people cannot discern the intervention and how it is used to change their behaviour – the transparency argument. To do so, the current study investigated public attitude towards four factors: ease of identification, willingness to change behaviour, perceived effectiveness, and acceptability; and their dependency on the transparency of an intervention, the agent that proposed the intervention and the rationale presented for the psychological method. In the main, participants were more accurate at identifying transparent compared to non-transparent interventions across the five contexts. In line with this, participants' judgement with regard to ease of identification, perceived effectiveness, and acceptability of an intervention, and their willingness to change their behaviour were consistently higher in transparent condition compared to the non-transparent condition. The four judgements were also perceived more favourably when a positive rationale was presented than when a negative rationale was presented. Overall, the findings supported all the hypotheses except hypothesis 6 where there were no significant differences in judgement with respect to the type of agent that proposed the nudges. Regression analysis revealed that non-transparent interventions decreased the responses for all judgements, and the presentation of positive rationales for interventions increased the positive direction of responses for all four judgements. As the age of the samples increased, judgements with regard to their willingness to change behaviour, perceived effectiveness and acceptability of the intervention decreased.

Also, when looking at regression analyses, interventions proposed by Government officials decreased judgments for willingness to change behaviour and acceptability; and judgements of willingness to change and perceived effectiveness decreased when the intervention was proposed by advertisers. Irrespective of transparency and rationale provided for the intervention, the presence of nudges in banking contexts were consistently judged as less favourable compared to the other four contexts. There was a sample differences for judgements of ease of identification and acceptability where US participants judged these less favourably than their UK counterparts. Lastly, acceptability of the intervention increased as the intervention become easier to identify, perceived as more effective, and when people were more willing to change their behaviour.

CHAPTER 7: General Discussion

7.1. Summary of Thesis Objectives

Nudging emerged from a branch of economics known as behavioural economics. They claim that, unlike econs, homo sapiens have bounded rationality so they sometimes make decisions that are not in their best interest. This is due to our limited cognitive capacities in processing knowledge and information. Therefore, if people are given sufficiently good information and feedback, they can make good decisions. According to nudge theory, our mind can be understood relative to the environment in which decisions are made. For example, people have a tendency to inertia and procrastination so implementing default rules can encourage people to enrol into pension schemes and register as organ donors. The central tenet of nudge is its philosophical framework “libertarian paternalism”, which states that it is legitimate for choice architects/policy makers to influence people’s behaviour in a welfare-promoting direction without restricting people’s freedom of choice. The concept has since attracted governments worldwide to incorporate nudge as part of social policy tools. To institutionalize nudge, an integral part is to ensure empirical testing, evidence, and transparency (Sunstein, 2014a); along with the ethical justifications to support its implementation in public policy making. Amongst these is a need for a sound theoretical framework for evaluating the nudge intervention and ensuring robust research design.

As discussed in Chapter 1 and 2, there is a lack of evidence to indicate the reliable effectiveness of nudges and another area that one requires closer inspection is the theoretical description of the mechanisms that support the way nudges are purported to work. The first objective was to examine the critical issues concerning the Dual System Theory (DST) and propose an alternative theoretical framework to re-consider the evidence base in the health domain, as a case in point. In addition, the thesis explored the ethical issues that arise from the rationale on which nudge is proposed to work, and the extent to which an endeavour that has global appeal should be a cause for concern. The second objective of the thesis was to empirically examine the extent to which nudge undermines autonomy: 1) the welfarist argument and 2) the transparency argument. From a welfarist perspective, if nudges such as defaults do not promote reflective thinking in any way, then they may risk promoting choices that are not aligned with the chooser’s higher-order desire. To examine the welfare-consequence of implementing defaults, the thesis included two empirical studies which investigated the extent to which people’s judgements of an individual’s true preference are sensitive to the implementation of

pro-self (i.e., maximise individual welfare) and pro-social (i.e., maximise social welfare) nudges in the context of retirement savings enrolment and organ donation registration respectively. On transparency grounds, an intervention that lacks transparency is typically regarded as manipulative. More so, influencing people's behaviour in a covert manner is considered ethically unacceptable. As with any public interventions, communicating policy decisions and the rationale behind them in an open and transparent way has been identified as one of the critical factors in increasing public acceptance (National Advisory Committee on Bioethics, 2015). Hence to successfully select and implement nudges, policy makers need to have an understanding of how nudges are perceived. Previous survey findings have shown that people generally prefer transparent nudges over non-transparent nudges, however, none of these studies have explicitly manipulated transparency to see how it influences people's judgement. This motivated the third empirical study to examine four factors that might influence public attitude towards nudge: ease of identification, willingness to change behaviour, perceived effectiveness and acceptability; and the dependency of these four factors on the transparency of interventions, presentation of rationales for the interventions and the agent that proposed the interventions. The main focus of this chapter is to provide an overview of the findings in relation to each research objectives proposed in this thesis and discuss the implication of these findings for informing public policy making.

7.2. Research Objective 1: Theoretical and Empirical Re-consideration

The first objective of this thesis was to provide a simple way of conceptualizing nudges that accounts for its evidence base in order to critically evaluate its effectiveness and any subsequent ethical justifications. In general, Type 1 nudges appeared to be short-lived and ineffective whereas Type 2 nudges seem to be more effective in reducing poor health behaviours such as alcohol consumption and cigarette smoking. In terms of survey findings, Type 1 nudges (e.g., default rules) were judged as least revealing people's preference because they do not deliberate over choices in the same way as active choice set-up. They were also perceived as less transparent compared to Type 2 nudges. These findings were in line with the proposed single system framework such that Type 1 nudges do not engage the decision maker on any substantial level to re-examine the basis on which the decisions are made, whereas Type 2 interventions encourage re-evaluation of behaviour to maintain coherence between information and behaviour change. The theoretical foundations of nudge, and the evidence-

base examining the efficacy of nudges in the health domain, therefore, suggested that there are good reasons to focus on implementing Type 2 over Type 1 nudges. In other words, there are theoretical grounds on which Type 2 nudges can be argued to have a more sustainable and deeper impact on generating positive behavioural change, and in line with this, the current evidence-base indicates, to some degree, that they are more effective than Type 1 nudges. Type 1, on the other hand, lacks empirical evidence to suggest that they work, and if they do, there is no good evidence to suggest that they are influencing behaviour on a non-conscious level (Osman, 2014a). From the outlook, any meaningful behaviour change in the long term (i.e. over a period of one year) is likely to result from the ability of the chooser to re-evaluate information from the choice context and their behaviour, and to do so requires making the chooser aware of the intention behind which the behaviour change is pursued.

7.3. Research Objective 2: The Welfarist Argument

The second objective of the thesis was evaluated in two parts. In part one, the thesis empirically examined the extent to which nudge is an intrusion on autonomy if nudges do not reflect chooser's higher-order desire ("true" preference) – the welfarist argument. The central claim of the nudge concept is *"to influence choice in a way that will make choosers better off, as judged by themselves"* (Thaler & Sunstein, 2008, p. 5). The emphasis here is the clause *"better off, as judged by themselves"* which explicitly states that the interpretation of one's welfare should be based on the individual's judgement rather than that of policy makers or other choice architects. This is also what underlies nudge's philosophical framework "libertarian paternalism". Within this, the choice architect has the right to reconstruct the choice architecture to reflect the individual's true preference. The crux of this lies in the interpretation of true preference, but these are hardly easy to identify. This is where the choice architect faces knowledge problems. In other words, the choice architect does not possess information about one's higher order preferences ("true" preference). Given that the choice architect can only go as far as observing the choices that people make,⁷ he/she will inevitably determine the chooser's true preference based on some general presumption of what the majority would choose and/or design the choice architecture based on his/her own value. On this view, the choice architect holds strong paternalistic views of his/her own about which available options

⁷ Sunstein and Thaler (2003) argue that people lack clear, stable, well-defined preferences in many domains and observing the choices they make does not reflect what they genuinely prefer.

will make decision-makers better off (Goldlin, 2015). For instance, if choice architects believe that people should eat healthier food, they will design the cafeteria to reflect that goal. This kind of value substitution can potentially undermine what people genuinely want and hence threaten their welfare. Chapter 4 and 5 presented the findings for third-party judgements of an individual's true preference to enrol onto retirement savings scheme and organ donation registers respectively. These two contexts have received much attention in the literature and represents major policy interventions. Even though these defaults promoted different ends (pro-self vs pro-social), the results seem to converge on one core finding. The key finding from both studies supported the general hypothesis that people's perceived true preference to register in a retirement savings scheme or donate their organs were perceived as higher in active choice systems (e.g., default opt-in and mandated choice) compared to passive choice system (default opt-out). These findings challenge the claim that nudges make people better off as judged by themselves. Even though the findings do not speak to issues regarding efficacy of different defaults in the pension enrolment contexts or the organ donation context, they do however address the issue with respect to the overall goodness of the consequence of a default. In the following section, I will discuss whether pro-self and pro-social can be best conceived as maximizing welfare and the implication of these findings.

Pro-self Default. As mentioned, the aim of the present study was to investigate how the public perceives, the preservation of an investor's free choice given their registration in different enrolment systems. If, as indicated in the present findings, choices made via a default opt-in/mandated choice system in the case of UK or a default opt-in system in the case of HK are perceived as more indicative of an investor's true preference to register in a retirement savings scheme, then it suggests that the public lean towards a form of active choice as signalling true preference. This would, in turn, indicate that the preservation of choice is important (Chapman, 2014; Osman, 2014a, 2016). In line with this finding, Carroll et al. (2009) contrasted mandated choice, default opt-out and default opt-in enrolments in savings plan and found that mandated choice was optimal (when participants procrastinate and/or have heterogeneous preference). The stickiness of defaults supports the concern that employees delay opting out of unsuitable defaults; it typically takes more than two years for the median employees to opt-out of with a savings rate of 2 or 3%. Similarly, Brown, Farrell, and Weisbenner (2011) found that individuals who are passively defaulted are substantially more likely to regret their decisions than those who made an active choice. This pattern of findings is also consistent with studies

examining preferences for automatic enrolment on organ donation registers (Johnson & Goldstein, 2003; Lin, Osman, Harris, & Read, 2018). Here too, the evidence suggests that preservation of free-choice is important and, critically, the implementation of a default opt-out system may lead to greater misclassification of individuals' preferences compared to a default opt-in system or a mandated choice system (Johnson & Goldstein, 2003). At a policy level, in line with these empirical findings, the Financial Conduct Authority (FCA) proposed that the choice architecture should enable consumers to make an active decision rather than offer 'a default' (Wells, 2014). However, this returns us back to the problem of people requiring considerable 'extra' help in navigating the information required to make a sensible long-term financial decision (Wells, 2014). Though from an ethical perspective, mandating people to choose where to invest their savings for retirement is a stronger alternative to the default opt-out system provided people are financially literate enough to make this kind of decision. It allows the investor to make a nuanced decision that fits their own needs and avoids the one-size-fits-all problem of defaults.

Pro-social Default. The findings show that when participants know that an individual has registered their decision to donate through some overt signal (i.e., under a mandated choice or a default opt-in system) this is likely perceived as a less ambiguous signal of a preference to donate. The signal generated from an active decision process typically encourages the agent to explicitly express a positive statement of intent, that is, request to donate or objection to donate. In line with this, there is evidence to suggest that people believe that the best way to obtain consent is for each individual to decide for themselves rather than leaving this decision to the family (Spital, 1993, 1995, 1996). Indeed, the study shows that a large majority of participants indicated that the deceased's wishes should be respected no matter what the family thinks. From the relatives' perspective, an active decision is easier to infer what the deceased would have wanted because the deceased recorded a positive wish to donate or objection to donate, as opposed to no objections to donate in the case of the default opt-out system. Ultimately, the question to ask is whether presumed consent can increase the number of people on the ODR, and at the same time represent a good signal of preference so that it maintains a high family consent rate as well. Based on the evidence from this study, it appears that presumed consent acts as a weak signal of true preference to donate, and may lead to sustained or higher family refusal rates than active choice to donate.

Although defaults are powerful solutions for decreasing intention-behaviour gaps in retirement savings and organ donation, the nudgers (choice architects – policy makers) can only go as far as constructing the choice architecture to increase the number of people registered on the ODR; because as aforementioned, they do not have sufficient information about the chooser's true preference. Sunstein (2015c) argue that realistically if the cost of opt-out is low, and if publicity and transparency are guaranteed, then there is far less threat to autonomy and welfare. However, transparency does not appear to be the case, at least not in the context of organ donation. The findings from Chapter 5 show that fewer than 50% of the participants in the default opt-in countries and only 19% from the default opt-out countries correctly identified the organ donation legislative system in their country, despite a recent survey showing around 80% of the population support organ donation 'in principle' (Department of Health & Social Care, 2017). Though this cannot be taken as a direct comparison to the findings presented in Chapter 5, and given the selected samples recruited from a different sampling technique (i.e. Mturk and Prolific Academic), the implication of this finding should be read with care. Nonetheless, this illustrates a potential issue with the public's lack of awareness of organ donation systems being implemented in their country, in particular, in presumed consent countries. This may help explain the weaker signal of perceived true preference to donate under a default opt-out system which was found in the study presented in Chapter 5. If this is the case, then families acting as proxy consent are likely to be unaware that a default opt-out system has been implemented and are equally likely to believe that the deceased were unaware of a default opt-out system. Hence the reason the deceased is registered on the ODR is because he/she may have forgotten to record an objection to donate during his/her lifetime. By inference, families are more likely to veto donation because, as revealed by our findings, signals attached to a default opt-out system suggest a weaker perceived underlying preference to donate; because a passive choice was made as opposed to an active one. This in part illustrates Rebonato's (2013) concern that even though both Germans (opt-in country) and Austrians (opt-out country) can reverse the default at low cost, the fact that almost 100% of Austrians and only 12% of Germans are organ donors suggest that in practice, they do not opt out for cognitive reasons (i.e. status quo bias, decisional inertia). This questions Thaler and Sunstein's (2003) claim that fewer opt-out means more people are satisfied with the default. The problem is that those who opt out are those that overcome inertia but this does not say much about the preferences of those that do not opt out. In Rebonato's view, the Austrians were given *nominal* freedom of choice but not *effective* freedom of choice.

On Welfare Grounds: Passive Choosing vs Active Choosing

Sunstein (2016c) claim that a good default is hard to reject on welfarist grounds. If we prefer active choosing, then we might as well rely on the theory that individuals usually know what is best for them. However, these findings suggest the welfare consequences of implementing either pro-self or pro-social defaults is questionable. The implementation of pro-social defaults in the organ donation context suggest that when someone has been defaulted into being an organ donor, his/her family member might hesitate before honouring the relevant choices for the reasons that both responsibility and intentions are ambiguous (Sunstein, 2015a). That is, they worry that the chooser's decisions might be a result of inertia and inattention. In the retirement savings context, the stickiness of default savings rate means that many employees delay opting out of defaults that are unsuitable for them. Certainly, if the barrier to opting out is high and if the default is not salient (i.e. inability to identify how and why an intervention is implemented), then it can potentially be consequential for matters that have a high stake (such as organ donation and retirement savings).

To implement welfare-promoting choice architectures, an alternative option is to adopt active choosing. This could eliminate a one-size-fits-all problem because it attends to a heterogeneous population and handles the issue of changing preferences and values over time (Sunstein, 2015a). Active choosing also encourages learning about a problem and promote the development of preferences. This has important implications as the learning process allows the strengthening of our “muscle” for future choice making (Sunstein, 2015a), for instance, while the invention of calculators meant that we save time on solving arithmetic, it has weakened our “muscle” for mental arithmetic over time. Similarly, if choices are made on behalf of us to choosing retirement savings plans, it simultaneously weakens our “muscle” for making autonomous decisions in the future. People whom made an active choice are also more likely to become invested in it. In areas such as organ donation where families act as proxy consent, active choosing offers distinctive signals (the deceased's intended decision as unambiguous preference) and has a distinctive meaning (Sunstein, 2015a). For these reasons, active choosing is believed to be welfare-enhancing. Therefore, it makes sense to ensure that the system which records an individual's choice to donate signals a genuine preference to donate. In the context of retirement savings, Carroll et al. (2009) found evidence that mandated choice increased initial enrolment by 28% compared to standard enrolment methods, producing a savings distribution three months after hire that would instead take 30 months to achieve under default

opt-in enrolment. In other domains, active choosing was found to have a larger effect in promoting green energy use than did green energy defaults (Hedlin & Sunstein, 2016). Admittedly, there are also drawbacks to active choosing because they can be cognitively effortful (including emotional effort and mental effort). For those who do not want to choose due to a lack information or experience, forcing to choose becomes a burden and may decrease welfare (Sunstein, 2015a). Indeed, we cannot possibly attend to every little detail in life, we simply do not have the cognitive capacities to do so. It is a fiction to be autonomous about everything in life, but for stakes that are high, there is a prima facie reason to take control. So, *what is the “right balance” of choosing that would promote autonomy and welfare?*

Choosing How to Choose

Given the nature of these findings, I do not argue against the use of default nor do I suggest that active choosing is necessarily the best choice architecture for everything in life. Having established the powerful effect of defaults and their lack of saliency, they should be used cautiously, especially when the choice architect does not have perfect information about people’s true preference.⁸ The current findings suggest that making a choice on their behalf without their awareness is likely to threaten their welfare. Whilst in the short-run at $t+0$, defaults may increase welfare as the cost of making a decision is reduced, in the long-run at $t+1$, it relieves an individual from learning about complex and important life decisions which are likely to be costly in the long-term (Binder & Lades, 2014). In this view, defaults would be less suitable if people have a defined preference or when learning matters (Sunstein, 2015a). Neither should they be adopted for situations where the stakes are high (e.g., contexts like organ donation, retirement savings). For example, the ends of not choosing your computer settings or adding a radio to your car are trivial whereas defaults implemented in the matters of life and death, or those that could impact one’s future savings for retirement can have consequences for one’s future welfare. But what is the right choice architecture to govern one’s welfare? In other words, what is the right balance between default and active choosing. I agree with Sunstein that defaults are unavoidable, after all, they are part of a choice architecture (Sunstein, 2015c). So perhaps, on welfare grounds, instead of choosing not to choose, we

⁸ A crucial goal is to determine people’s true preference and to ensure that nudges do in fact improve people’s welfare by influencing their behaviour in line with their true preference. This can be made possible by adopting preference identification methods such as revelatory frame for a more detailed discussion on this, see Goldlin (2015).

should promote choosing *how* to choose along with providing information and education. My view is therefore in line with Appelt, Gao, Johnson, and von Glahn's (2014): "*Giving people a choice of how to choose preserves their autonomy and helps them make better choices*". For those that enjoy the intrinsic value of making a choice, they will choose to choose by taking matters in their own hands, and for those who do not enjoy choosing, then there is the option to opt into a default that takes care of things for them. In a way, this works similarly to direct debits in the commercial context, where people can either set up automatic monthly payment or opt to manually make a payment quarterly. This approach overcomes the common claim that people have inconsistent preferences (Sunstein & Thaler, 2003; Thaler & Sunstein, 2008) or the claim that individuals are in the best position to know what is best for them (Mill, 1863) because choosing how to choose is in itself a learning process about one's own preference.

However, it is worth noting that this proposition may not work so well with the organ donation context because the decision in these contexts are one-off and there isn't any corrective measure if one forgets to make a choice. Whatever choice one makes during his/her lifetime (whether actively opt-in, default opt-out, or made no choice at all) and the signals inferred from these systems will have an influence on the relative's decisions to consent or veto. If the goal of the policy maker is to increase the number of actual donation by encouraging those who genuinely want to donate to register on the ODR, then I recommend that efforts in this area should focus on information framing which will be more likely to reflect people's true preference (or in other words, to avoid ambiguous signal being perceived by relatives). For instance, the Behavioural Insights Team (BIT) conducted one of the largest randomized controlled trials (1,085,322 individuals) ever run in the UK testing the effect of including different messages on a high traffic website on GOV.UK that encourages people to actively join the NHS ODR. The results showed that the best performing message was "if you needed an organ transplant, would you have one? If so please help others." and the least performing message was "Every day thousands of people who see this page decide to register" which contained a picture (Behavioural Insights Team, 2013). The best performing message by itself lead to over 350,000 registrations for organ donation via the GOV.UK link (Loosemore, 2014). These findings suggest that a simple change of message can make a vast difference to the number of people actively signing up to the register. This type of online trial would also be more cost effective and less risky compared to implementing a default opt-out system, and

would be more likely to increase family consent rate because the strength of true preference to donate is stronger when an active choice is made.

Overall, the welfarist argument for nudges such as default rules is that if transparency and cost of opt-out cannot be guaranteed then the policy maker must be able to show evidence that the nudge implemented allows some level of deliberative action so as to avoid value substitution. Otherwise, it seems the better alternative is to leave it to people to choose how they want to choose as a default that lacks saliency could project different perceived preference. Last but not least, autonomy and welfare go *hand in hand*. If people can identify the how and the why of an intervention then they have the autonomous power to act according to their true preference. For the policy maker, this means they are also less likely to impose value substitution as the threat to welfare is minimized by token transparency.

7.4. Research Objective 2: The Transparency Argument

The second objective of the thesis was evaluated in two parts. In part two, the thesis empirically examined how autonomy is best preserved from the perspective of transparency. A central tenet in the criticism of nudges states that interventions lacking transparency are regarded as manipulation of choice. Certainly, if people have not consented to manipulative interventions, nudges can undermine both autonomy and dignity which can be ethically objectionable (Sunstein, 2015c). For people to consent to manipulative intervention, people must be able to detect the intervention in the first place. If nudges are to become a social policy instrument, then influencing people without their awareness can potentially threaten their autonomy and welfare. Sunstein (2014a) argue that although nudges are considerably less overt than traditional policy measure, it should “*never take forms of manipulation or trickery. The public should be able to review and scrutinize nudges no less than government actions of any other kinds*”. However, the covertness of nudge raises questions as to whether it provides sufficient transparency and accountability (Baldwin, Cave, & Lodge, 2011). For example, if a nudge bypasses people’s cognitive capacity for reasoning to make a choice other than what they would normally have chosen, then this leads to questions around the government’s intentions. Given the mechanism by which nudges works (e.g., Type 1 nudges), it would seem difficult for people to be able to review and scrutinize a nudge even if they want to. In this view, transparency is the guiding principle for most ethical justifications of nudges, with the foremost being autonomy. However, it appears that much of the debate concerning autonomy has largely

proceeded without taking into account public attitudes towards a diverse range of dimensions associated with nudges (i.e. their efficacy, their acceptability, the ease with which the underlying basis for behavioral change can be identified, and one's willingness to change as a result of a nudge being implemented). Public attitudes towards nudge interventions are likely to be a key barriers for facilitating policy implementations (Hagman et al., 2015). This motivated the aim of the third empirical study to examine the extent to which public attitudes are dependent on the transparency of interventions. To do so, the thesis provided a notion of transparency from the perspective of those being nudged. This is because the definition of transparency to date has been defined in relation to the intention of the choice architect/policy maker, though this is important, what really matters is whether those being nudged can actually identify the intervention being implemented. Subsequently, this definition was used to assess public attitude towards transparency of interventions in relation to four factors: ease of identification, perceived effectiveness and acceptability, and people's willingness to change their behaviour under transparent and non-transparent conditions. In addition, the study also investigated the extent to which these judgements are dependent on the expert that proposed the intervention and the rationale presented for the intervention. Below I will discuss these four judgements in turn.

Transparency. To distinguish nudge from its manipulative use, the intervention must be transparent enough so that the chooser can discern the intention and means of the behaviour change being pursued (Hansen & Jespersen, 2013). Thus, an increase in transparency will enable the chooser to recognize how and why an intervention has been implemented (Bovens, 2009). In the main, the findings from Chapter 6 suggested that across all five contexts, participants found it easier to identify the intervention and how it changes their behaviour when the intervention is transparent as opposed to non-transparent. In line with this finding, over 70% of the participants were generally accurate at identifying transparent nudges compared to only 50% correctly identifying non-transparent nudges. This shows that people's assessment of transparency is somewhat similar to those of the researcher's for transparent interventions but the notion of transparency becomes less clear in the non-transparent conditions. In this regard, it is worth considering the fact that public surveys of nudges also suggest that the public show much higher approval ratings for transparent over non-transparent nudges (Arad & Rubinstein, 2015; Felsen et al., 2013; Hagman et al., 2015; Hedlin & Sunstein, 2016; Jung & Mellers, 2016; Mazzocchi et al., 2015; Reisch & Sunstein, 2016; Reisch, Sunstein, & Gwozdz, 2016; Sunstein,

2016c; Sunstein, Reisch, & Rauber, 2017). Again, this goes to show that, whether or not people are going to modify their behaviour in light of nudges, or more typical policy methods of behavioural change, they are supportive of explicit methods that signal what methods are being used and how they change behaviour over those that seek to do this covertly, especially without their consent (Osman, 2016). In line with this, most recently, Sunstein, Reisch, & Rauber (2017) showed that many people do care about freedom of choice and they will reject many well-motivated policies that do not allow for that kind of freedom.

Willingness to Change Behaviour. The findings subsequently showed that people were more willing to change their behaviour when the intervention was transparent as opposed to non-transparent. In other words, their propensity to identify with the goals of the intervention increases when the intention and means of the behaviour change being pursued is clear. This also suggests that transparent interventions were more likely to be in line with their first-order desire (i.e., “true” preference) either to eat healthier, smoke less, or to save more, and so on. This is important because without sufficient motivation or the “*brain processes that energize and direct behaviour*” (Michie, van Stralen, & West, 2011), behaviour change is less likely to succeed. Furthermore, it can be suspected that people’s willingness to change their behaviour is also dependent on the objective of the nudge. Reisch & Sunstein (2016) found that people’s judgements are usually dependent on whether a nudge is well-motivated and leads to good consequences (from the standpoint of most people whom they affect). In the case of transparent nudges, the intention and means with respect to improving health and welfare are comprehensible to the chooser. This is an important contributing factor in terms of the individual’s objective and capacity to influence their own behaviour. Nudging in this manner makes it easier to adopt and carry out behaviour change as a habit, and eventually be integrated into the social environment that further endorse the behaviour (Junghans, Cheung, & De Ridder, 2015).

Perceived Effectiveness. There is concern that making nudges transparent reduces their effectiveness (Bovens, 2009; House of Lords, 2011) while others argue that revealing a nudge may cause people to protest against the intervention (Arad & Rubinstein, 2015). The findings suggest otherwise: transparent interventions were perceived as more effective than non-transparent interventions. Junghans, Cheung, & De Ridder (2015) found that what made nudges potentially effective was that they subtly facilitated the targeted behaviours. According

to the authors, the intervention is more likely to be perceived as effective if it made behaviours easier or more fun to perform, and made the choices more salient. In line with this, Loewenstein et al. (2015) found that informing people about the use of defaults for decisions about advance directive did not lower their effectiveness. Similarly, Bruns, Kantorowicz-Reznichenko, Klement, Jonsson, & Rahali (2016) found that nudges in the form of defaults to increase contributions to carbon emission can be transparent, and effective. These empirical findings so far have all suggest that nudges can be transparent without compromising effectiveness.

Public Acceptability In line with ease of identification, perceived effectiveness, and willingness to change behaviour, people also found transparent interventions more acceptable than non-transparent interventions. This conforms with previous evidence showing that overt interventions are perceived as more acceptable than covert interventions (Felsen et al., 2013). Hence, people tend to prefer transparent nudges that target deliberative process than non-transparent nudges that target non-conscious processing (Sunstein, 2016b). Although Hall et al. (2018) found that this was not the case for pictorial warnings on cigarette packaging (non-transparent) as these were perceived as more acceptable, they did not examine support for text-only warnings on packs so it would be difficult to see how the outcomes would have differed if they included text-only warnings as well. Finally, regression analyses showed that all three judgements increased public acceptability which are in the right direction. This shows that people will find the intervention more acceptable as people find an intervention easier to identify, perceived it as more effective, and are more willing to change their behaviour. In line with this, Petrescu et al. (2016) found that the public will be more likely to accept an intervention if it is perceived as more effective. Similarly, Pechey, Burge, Mentzakis, Suchrcke, and Marteau (2014) found that both the type of intervention and its anticipated effectiveness influenced public acceptability for interventions designed to decrease alcohol consumption. By far, this is good news for nudge because a policy without public support does not stand its ground. Similarly, it means it will be easier for policy makers to publicly defend a policy if the public find it acceptable.

Rationale for the Intervention. Intuitively, to perceive an intervention as effective requires knowing its psychological mechanism, in other words the psychological basis on which it is designed to work. Bovens (2009) concluded that revealing how an intervention works is also part of token transparency. The current findings suggest that presenting positive rather than

negative rationale for the psychological mechanism behind an intervention increased all four judgements. For judgements regarding willingness to change and public acceptability, presentation of positive rationales was perceived more favourably compared to presentation of positive & negative rationales, though there were no difference between these rationales for judgements of ease of identification and perceived effectiveness. To my knowledge, there has not been any literature on the effect of providing rationale on people's judgements. These findings provide insights into Sunstein's (2016c) and Bovens (2009) speculation as to whether disclosing the psychological mechanism of nudge might increase its impact. By revealing how an intervention positively works to change people's behaviour, people will find it easier to identify the intervention, perceived it as more effective and acceptable, and would be more willing to change their behaviour. Therefore, the results would have implications as to whether government action is justifiable, because Sunstein (2016c) argues that government action requires the provision of reasons for why the intervention is being introduced. Consistent with the current finding, a qualitative interview study of UK consumers found that most consumers approved of the concept of nudge if it was explained to them, especially in the health domain, provided that nudges are designed to benefit individuals and society; and that consumer comprehended the decision-making context and the reasoning behind the promotion of the targeted behaviour (Junghans et al., 2015).

Lastly, across all four judgements, the banking context were consistently judged less favourably than the health contexts. But, perhaps, the largest difference between banking and these other four contexts was the judgement concerning acceptability. Most recently, Sunstein, Reisch, & Reuber (2017) conducted a global survey examining consensus of nudging, and found that most citizens are enthusiastic about nudge so long as they are consistent with people's values and interests. One reason for the difference in findings between the banking and the health context is likely to have been due to the banking nudge intervention being perceived as less consistent with people's interest or values. Though it is also possible that people simply do not want choice architects to interfere with their finances. Intuitively, one would not trust either the government, researcher or an advertising agent with her money at least not in the context set out in this study, but without further empirical testing these interpretations should be read with care.

Furthermore, the study found that people's willingness to change their behaviour and their acceptability of nudges were perceived less favourably when the interventions were proposed by government officials. Similarly, both willingness to change and perceived effectiveness were judged less favourably when nudges were proposed by advertisers. The reason for this is possibly because the psychological methods are implemented in contexts in which the two agents (Government, Advertiser) are not seen as experts in the field. Junghans, Cheung, & De Ridder (2015) found that nudges are more highly approved when they are implemented by experts and industry as opposed to policy makers. More specifically, they would have greater trust in actors who are credible such as specialised expertise in a specific subject, for example, for nudges that target particularly at healthy eating, health experts may be seen as more credible than policy makers.

Taken together, these results are consistent with previous findings that a person's autonomy is better preserved when an intervention is transparent as opposed to non-transparent. In line with Jung and Mellers' (2016) findings that System 2 (transparent) nudges were viewed as effective for better decision making and more necessary for changing behaviour, whereas System 1 (non-transparent) nudges were perceived as more autonomy-threatening. Similarly, Sunstein (2016c) argue that System 2 nudges are better on welfare grounds and will best promote autonomy. Certainly, if people can detect how their behaviour is being changed then they can either accept or reject a nudge on the basis of whether it aligns with their true preference. It also becomes possible for people to acknowledge a potential problem caused by their current suboptimal decisions, for example, to think about whether they should be saving for their retirement. A non-transparent nudge, on the other hand, means that it is difficult for people to rationalise how it might change their behaviour positively. Transparent nudges therefore allow individuals to act reflectively so that their propensity to change is aligned with the goal of the intervention. In other words, transparent nudges respect the chooser's autonomy because under these conditions the chooser can be said to have control over his/her deliberative capacity to act in accordance with his/her understanding of welfare. Last but not least, transparent interventions are more likely to be perceived as effective and acceptable, and people are more willing to change their behaviour under these conditions. These findings provide insights for policy making and provide a strong rationale for choosing transparent over non-transparent nudges.

7.5.A Lesson on Autonomy

The empirical findings from this thesis converges to one view, that is, transparency is the cornerstone to preserving autonomy and promoting welfare. This view is in line with Sunstein (2016c) who states that transparency is a necessary condition for both nudge and choice architecture. From a welfarist perspective, a lack of transparency means people are being defaulted into a programme that may not align with their higher-order desire and this is consequential for matters that have a high stake (e.g., in organ donation context and pension enrolment). In line with this, the public will generally favour transparent nudges because it is easier to identify the interventions, and how their behaviour are changed. The public will also perceive an intervention as more effective and more acceptable, and more likely to reflect their values and interests (i.e., more willing to change their behaviour because they actually want to change their behaviour in that context). If people are manipulated into making choices without their awareness then naturally they are also relieved of the responsibility associated with these choices. Rather we should hold people responsible for the choice they make even if it means that they do not make the optimal choices, but nonetheless they are more likely to make choices that reflect their true preferences without introducing moral hazard in the form of intruding their autonomy. Additionally, choice making is a learning process during which people come to an understanding about the possible effects of their decision-making, and it enables them, if they are willing and able to integrate new knowledge that they encounter, to implement corrective measures in future decisions. As explained by Hausman and Welch (2010): “*the risk of exploiting decision-making foibles will ultimately diminish people’s autonomous decision-making capacities*”, they becomes less willing to invest the effort in making decisions. For example, this would mean, through learning, enabling a transition of eating less in a restaurant to eating healthy at home. Therefore, holding people more responsible for their choices encourages both learning by education and learning by error.

7.6. Conclusion

The choice architecture forms the basis of our everyday decision making and the history of judgment and decision-making research in psychology suggests that we are susceptible to cognitive biases, and in turn we often make suboptimal decisions. Most of us value our health yet act in ways that undermines it. Nudges have been proposed as a tool to address this gap between value and behaviour by rearranging the choice context. But because nudges mainly

work by exploiting our cognitive biases, there are reasons to believe they are ethically more worrisome than traditional policy instruments which are considered more explicit. Since nudges bypass reflective thinking, it is said to *work best in the dark* (i.e. non-transparent to the decision-maker). The inability to detect how and why the intervention was implemented therefore could infringe upon autonomy and welfare. But since the literature on nudge has been rooted in different theories of agency that are incompatible with the nudge approach, the thesis proposed an alternative framework to account for its evidence base before evaluating where the current ethical concerns stand. The more we understand how nudges work, the better armed we (social scientists, policy makers, practitioner) are in designing ways of intervening on behaviours to achieve the best outcome for individuals that need and want it and which are ethically justifiable on these grounds. Hence, any meaningful change in behaviour arises from developing a consistently coherent basis on which people understand the reasons for their decisions and how they enact them.

The proposed single system framework suggested that to encourage people to help themselves, for example, in targeting serious problems around NCDs, we need to make the goal of helping oneself making better lifestyle choices through a coherent and sustained approach. From a theoretical perspective, this is best achieved through Type 2 rather than Type 1 nudges. This proposition corroborated with the empirical findings on ethical concerns related to intrusion on autonomy. The evidence suggests that Type 1 nudges fair less well with respect to autonomy and welfare because Type 1 nudges do not engage the decision maker on any substantial level for the decision-maker to make autonomous decisions in line with his/her true preference. If nudge is to promote sustainable behaviour on a long-term basis, there needs to be a sufficient level of transparency so that the decision-maker can re-evaluate his/her behaviour to maintain the greatest coherence between information and behavioural change. To incorporate nudging as part of social policy tool, transparency should be built into practice. This includes being transparent and open about their intention so that their decisions are subject to public scrutiny and review, but equally important is, the intervention being implemented should be transparent to those being nudged. This applies to any democratic governments. On theoretical, empirical and ethical grounds, Type 2 nudges that work in a more transparent manner are easier to identify and are more likely to be perceived as effective and acceptable in promoting behaviour change that are in line with people's high-order desires (or "true" preference). In this view, transparency and scrutiny reduce the likelihood of welfare-reducing choice architectures.

Lastly, the private sector (e.g., advertising) inevitably nudge (in a non-transparent manner) with the goal of promoting suboptimal choices for commercial purposes. To combat these opposing forces, nudges should enable people to exercise their autonomy so that they can discern the intention of the nudge used to change their behaviour. This is because transparency will promote re-evaluation of their behaviour and their own values (e.g., acknowledging their suboptimal decisions) and to “learn” to resist potentially seductive options promoted through nudges from advertisers and other private-sector agents. Without transparency, it would be akin to two invisible forces attacking each other. Nudging, in a sense, acts like a catalyst to lower the “activation energy” (effort) in a “chemical reaction” (behaviour change). Because most suboptimal behaviours arise from limited self-control, a lack of relevant knowledge, and effort/motivation required to carry out the action as intended. Therefore, for nudge to lower the effort for *effective* behaviour change (here, effective refers to sustainable or habitual changes), it should be sufficiently transparent for people to exercise their autonomy in order for people to re-evaluate the information and behaviour for coherence. For example, to lose weight, one has to learn the nutritional value and calorie content in order to understand the content of a healthy balanced diet. Type 2 nudges like calorie labelling or traffic light labelling will enable an individual to re-evaluate that information in connection to their own values or interests, which in turn lowers the effort required to implement behavioural change (i.e., adopting a low-calorie meal). In closing, a prerequisite requirement for any liberal democratic society is to be transparent about the working principles of its policy, and to design interventions that enhance people’s autonomy and welfare. The thesis suggests that on theoretical, empirical, and ethical reasons, this is best achieved through Type 2 rather than Type 1 nudges.

REFERENCE

- Abadie, A., & Gay, S. (2006). The impact of presumed consent legislation on cadaveric organ donation: a cross-country study. *Journal of Health Economics*, 25(4), 599–620. <https://doi.org/10.1016/j.jhealeco.2006.01.003>
- Ameriks, J., Caplin, A., & Leahy, J. (2003). Wealth accumulation and the propensity to plan. *Quarterly Journal of Economics*, 118(3), 1007–1047.
- Andersen, R. E., Franckowiak, S. C., Snyder, J., Bartlett, S. J., & Fontaine, K. R. (1998). Can Inexpensive Signs Encourage the Use of Stairs? Results From a Community Intervention. *Annals of Internal Medicine*, 129(5), 363–369.
- Anthes, W. L. (2004). Financial Illiteracy in America: A Perfect Storm, a Perfect Opportunity. *Journal of Financial Service Professionals*, 58(6), 49–56.
- Appelt, K. C., Gao, J., Johnson, E. J., & von Glahn, T. (2014). Choosing How to Choose: Can People Choose the Best Choice Architecture? *Association for Consumer Research*, 42, 295–396.
- Arad, A., & Rubinstein, A. (2015). *The People's Perspective on Libertarian-Paternalistic Policies*.
- Ashcroft, R. E. (2013). Doing good by stealth: comments on “Salvaging the concept of nudge”. *Journal of Medical Ethics*, 39(8), 494. <https://doi.org/10.1136/medethics-2012-101109>
- Attwood, A. S., Scott-Samuel, N. E., Stothart, G., Munafò, M. R., & Campanella, S. (2012). Glass Shape Influences Consumption Rate for Alcoholic Beverages. *PLoS ONE*, 7(8), e43007. <https://doi.org/10.1371/journal.pone.0043007>
- Australian Government - Department of Health. (2016). *Post-Implementation Review Tobacco Plain Packaging 2016*. Retrieved from <http://ris.pmc.gov.au/sites/default/files/posts/2016/02/Tobacco-Plain-Packaging-PIR.pdf>
- Åvitsland, A., Solbraa, A. K., & Riiser, A. (2017). Promoting workplace stair climbing: sometimes, not interfering is the best. *Archives of Public Health*, 75, 2. <https://doi.org/10.1186/s13690-016-0170-8>
- Baer, J. S., Stacy, A., & Larimer, M. (1991). Biases in the perception of drinking norms among college students. *Journal of Studies on Alcohol*, 52(6), 580–586.
- Baldwin, R. (2014). From Regulation to Behaviour Change: Giving Nudge the Third Degree. *The Modern Law Review*, 77(6), 831–857. <https://doi.org/10.1111/1468-2230.12094>
- Baldwin, R., Cave, M., & Lodge, M. (2011). *Understanding Regulation: Theory, Strategy and Practice* (2nd ed.). Oxford: Oxford University Press.

- Barber, K., Falvey, S., Hamilton, C., Collett, D., & Rudge, C. (2006). Potential for organ donation in the United Kingdom: audit of intensive care records. *BMJ (Clinical Research Ed.)*, *332*(7550), 1124–1127. <https://doi.org/10.1136/bmj.38804.658183.55>
- Barton, A., & Grüne-Yanoff, T. (2015). From Libertarian Paternalism to Nudging—and Beyond. *Review of Philosophy and Psychology*, *6*(3), 341–359. <https://doi.org/10.1007/s13164-015-0268-x>
- Behavioural Insights Team. (2013). *Applying Behavioural Insights to Organ Donation: preliminary results from a randomised controlled trial*. Retrieved from <https://gds.blog.gov.uk/2014/03/18/organ-donor-register/>
- Benartzi, S., & Thaler, R. H. (2013). Behavioral Economics and the Retirement Savings Crisis. *Science*, *339*(6124), 1152–1153.
- Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). The Importance of default options for retirement saving outcomes: Evidence from the United States. In S. J. Kay & T. Sinha (Eds.), *Lessons from Pension Reform in the Americas* (pp. 59–87). Oxford: Oxford University Press.
- Binder, M., & Lades, L. K. (2014). *Autonomy-enhancing paternalism*.
- Blamey, A., Mutrie, N., & Aitchison, T. (1995). Health promotion by encouraged use of stairs. *BMJ*, *311*(7000), 289–290. <https://doi.org/10.1136/bmj.311.7000.289>
- Blumenthal-Barby, J. S., & Burroughs, H. (2012). Seeking Better Health Care Outcomes: The Ethics of Using the “Nudge.” *The American Journal of Bioethics*, *12*(2), 1–10. <https://doi.org/10.1080/15265161.2011.634481>
- Bonell, C., McKee, M., Fletcher, A., Wilkinson, P., & Haines, A. (2011). One nudge forward, two steps back. *BMJ*, *342*, d401–d401. <https://doi.org/10.1136/bmj.d401>
- Boreham, C. A. G., Wallace, W. F. M., & Nevill, A. (2000). Training Effects of Accumulated Daily Stair-Climbing Exercise in Previously Sedentary Young Women. *Preventive Medicine*, *30*(4), 277–281. <https://doi.org/10.1006/pmed.2000.0634>
- Borsari, B., & Carey, K. B. (2003). Descriptive and injunctive norms in college drinking: a meta-analytic integration. *Journal of Studies on Alcohol*, *64*(3), 331–341.
- Botti, S., & McGill, A. L. (2006). When Choosing Is Not Deciding: The Effect of Perceived Responsibility on Satisfaction. *Journal of Consumer Research*, *33*(2), 211–219. <https://doi.org/10.1086/506302>
- Bovens, L. (2009). The Ethics of Nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference Change : Approaches from Philosophy, Economics and Psychology* (pp. 217–

- 209). Dordrecht: Springer Netherlands.
- Bowels, S. (1998). Endogenous Preferences: The Cultural Consequences of Markets and other Economic Institutions. *Journal of Economic Literature*, *XXXVI*, 75–111.
- Branson, C., Duffy, B., Perry, C., & Wellings, D. (2012). *Acceptable behaviour? Public opinion on behaviour change policy*. London, UK.
- Brose, L. S., Chong, C. B., Aspinall, E., Michie, S., & McEwen, A. (2014). Effects of standardised cigarette packaging on craving, motivation to stop and perceptions of cigarettes and packs. *Psychology & Health*, *29*(7), 849–860. <https://doi.org/10.1080/08870446.2014.896915>
- Brown, J. R., Farrell, A. M., & Weisbenner, S. J. (2011). *The downside of defaults*. Retrieved from www.nber.org/programs/ag/rrc/NB11-01 Brown, Farrell, Weisbenner FINAL.pdf.
- Brownell, K. D., Stunkard, A. J., & Albaum, J. M. (1980). Evaluation and modification of exercise patterns in the natural environment. *The American Journal of Psychiatry*, *137*(12), 1540–1545.
- Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Jonsson, M. L., & Rahali, B. (2016). *Can Nudges Be Transparent and Yet Effective?* (No. WiSo-HH Working Paper Series No. 33).
- Bucher-Koenen, T., & Lusardi, A. (2011). Financial literacy and retirement planning in Germany. *Journal of Pension Economics and Finance*, *10*(4), 565–584.
- Bullers, S., Cooper, M. L., & Russell, M. (2001). Social network drinking and adult alcohol involvement. *Addictive Behaviors*, *26*(2), 181–199. [https://doi.org/10.1016/S0306-4603\(00\)00099-X](https://doi.org/10.1016/S0306-4603(00)00099-X)
- Burk, J., Huang, A. A., & Luoto, J. E. (2015). *Automatic Enrollment in Retirement Savings Vehicles - Evidence from the Health and Retirement Study* (No. WR-1117).
- Byrne, A., Blake, D., Cairns, A., & Dowd, K. (2007). Default funds in UK defined contribution pension plans. *Financial Analysts Journal*, *63*(4), 4748.
- Cabinet Office Behavioural Insights Team. (2011). *Behaviour Insights Team: Annual Update 2010-11*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60537/Behaviour-Change-Insight-Team-Annual-Update_acc.pdf
- Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2009). Optimal Defaults and Active Decisions. *QJEcon*, *124*(4), 1639–1674.
- Chapman, G. (2014). Judge, Nudge, Dodge: Doing What’s Good for Us. In *SJDM Presidential*

Address. Long Beach, CA.

- Choi, J. J., Laibson, D., & Madrian, B. C. (2004). Plan Design and 401(k) Savings Outcomes. *National Tax Journal*, *57*(2), 275–298.
- Cohen, I. G., Lynch, H. F., & Robertson, C. T. (2016). *Nudging Health: Health Law and Behavioral Economics*. (I. G. Cohen, H. F. Lynch, & C. T. Robertson, Eds.). Johns Hopkins University Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.
- Coleman, K. J., & Gonzalez, E. C. (2001). Promoting Stair Use in a US–Mexico Border Community. *American Journal of Public Health*, *91*(12), 2007–2009. <https://doi.org/10.2105/AJPH.91.12.2007>
- Davidai, S., Gilovich, T., & Ross, L. D. (2013). The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences*, *109*(38), 15201–15205. <https://doi.org/10.1073/pnas.1211695109>
- De Neys, W. D. (2006). Dual Processing in Reasoning: Two Systems but One Reasoner. *Psychological Science*, *17*(5), 428–433. <https://doi.org/10.1111/j.1467-9280.2006.01723.x>
- DeJong, W., Franz, H. G., Wolfe, S. M., Nathan, H., Payne, D., Reitsma, W., & Beasley, C. (1998). Requesting organ donation: an interview study of donor and nondonor families. *American Journal of Critical Care*, *7*(1), 13–23.
- Den Hartogh, G. (2012). The role of the relatives in opt-in systems of postmortal organ procurement. *Medline, Health Care, and Philosophy*, *15*(2), 195–205. [https://doi.org/DenHartogh, G. \(2012\). The role of the relatives in opt-in systems of postmortal organ procurement. Medicine, Health Care, and Philosophy, 15\(2\), 195–205. <http://doi.org/10.1007/s11019-011-9317-0>](https://doi.org/DenHartogh, G. (2012). The role of the relatives in opt-in systems of postmortal organ procurement. Medicine, Health Care, and Philosophy, 15(2), 195–205. http://doi.org/10.1007/s11019-011-9317-0)
- Department for Work & Pensions. (2008). Pension Act 2008. Retrieved from <http://www.legislation.gov.uk/ukpga/2008/30/contents>
- Department of Health & Social Care. (2017). *Consultation on introducing ‘opt-out’ consent for organ and tissue donation in England*. Retrieved from <https://www.gov.uk/government/consultations/introducing-opt-out-consent-for-organ-and-tissue-donation-in-england/consultation-on-introducing-opt-out-consent-for-organ-and-tissue-donation-in-england>
- Downs, J. S., Loewenstein, G., & Wisdom, J. (2009). Strategies for Promoting Healthier Food

- Choices. *American Economic Review*, 99(2), 159–164.
<https://doi.org/10.1257/aer.99.2.159>
- Dunlop, S. M., Dobbins, T., Young, J. M., Perez, D., & Currow, D. C. (2014). Impact of Australia's introduction of tobacco plain packs on adult smokers' pack-related perceptions and responses: results from a continuous tracking survey. *BMJ Open*, 4(12), e005836. <https://doi.org/10.1136/bmjopen-2014-005836>
- Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge, UK: Cambridge University Press.
- Dworkin, G. (2012). Lying and nudging. *Journal of Medical Ethics*, 39(8), 496–497.
- Elbel, B., Gyamfi, J., & Kersh, R. (2011). Child and adolescent fast-food choice and the influence of calorie labeling: a natural experiment. *International Journal of Obesity (London)*, 35(4), 493–500. <https://doi.org/10.1038/ijo.2011.4>
- Emrich, T. E., Qi, Y., Lou, W. Y., L'Abbe, M. R., Garcia, R., & Cavalcanti, R. (2017). Traffic-light labels could reduce population intakes of calories, total fat, saturated fat, and sodium. *PLOS ONE*, 12(2), e0171188. <https://doi.org/10.1371/journal.pone.0171188>
- Ennett, S. T., Bauman, K. E., Hussong, A., Faris, R., Foshee, V. A., Cai, L., & DuRant, R. H. (2006). The Peer Context of Adolescent Substance Use: Findings from Social Network Analysis. *Journal of Research on Adolescence*, 16(2), 159–186. <https://doi.org/10.1111/j.1532-7795.2006.00127.x>
- Epstein, S. (1973). The self-concept revisited: Or a theory of a theory. *American Psychologist*, 28(5), 404–416. <https://doi.org/10.1037/h0034679>
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *The American Psychologist*, 49, 729–724.
- Eurobarometer. (2007). *Europeans and organ donation report*.
- Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. *International Journal of Psychology* (Vol. 31). East Sussex: Psychology Press.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Exley, M., White, N., & Martin, J. H. (2002). Why families say no to organ donation. *Critical Care Nurse*, 22(6), 44–51.
- Fardy, P. S., & Ilmarinen, J. (1975). Evaluating the effects and feasibility of an at work stairclimbing intervention program for men. *Medicine and Science in Sports*, 7(2), 91–93.

- Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: public attitudes towards overt and covert nudges. *Judgement and Decision Making*, 8(3), 202. <https://doi.org/10.1177/2332858416674007>
- Felsen, G., & Reiner, P. B. (2015). What can neuroscience contribute to the debate over nudging? *Review of Philosophy and Psychology*, 6(3), 469–479. <https://doi.org/10.1007/s13164-015-0240-9>
- Financial Services and the Treasury Bureau. (2015). *Mandatory Provident Fund Schemes (Amendment) Bill 2015*. HK.
- Finkelstein, E. A., Strombotne, K. L., Chan, N. L., & Krieger, J. (2011). Mandatory Menu Labeling in One Fast-Food Chain in King County, Washington. *American Journal of Preventive Medicine*, 40(2), 122–127. <https://doi.org/10.1016/j.amepre.2010.10.019>
- Foxcraft, D. R., Moreira, M. T., Santimano, N. M. L., & Smith, L. A. (2015). Social norms information for alcohol misuse in university and college students (Review). *Cochrane Database of Systematic Reviews*, (12). <https://doi.org/10.1002/14651858.CD006748.pub4>
- Ghorbani, F., Khoddami-Vishteh, H. R., Ghobadi, O., Shafaghi, S., Louyeh, A. R., & Najafizadeh, K. (2011). Causes of family refusal for organ donation. *Transplantation Proceedings*, 43(2), 405–406. <https://doi.org/10.1016/j.transproceed.2011.01.031>
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4), 684–704. <https://doi.org/http://dx.doi.org/10.1037/0033-295X.102.4.684>
- Gigerenzer, G., & Regier, T. (1996). How Do We Tell an Association From a Rule? Comment on Sloman (1996). *Psychological Bulletin*, 119(1), 23–26.
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.
- Goldlin, J. (2015). Which Way To Nudge? Uncovering Preferences in the Behavioral Age. *The Yale Law Journal*, 125(1), 226–270.
- Gomberg, L., Schneider, S. K., & DeJong, W. (2001). Evaluation of a social norms marketing campaign to reduce high-risk drinking at the University of Mississippi. *The American Journal of Drug and Alcohol Abuse: Encompassing All Addictive Disorders*, 27(2), 375–389. <https://doi.org/10.1081/ADA-100103715>
- Gomberg, L., Schneider, S. K., & DeJong, W. (2001). Evaluation of a social norms marketing

- campaign to reduce high-risk drinking at the University of Mississippi*. *The American Journal of Drug and Alcohol Abuse: Encompassing All Addictive Disorders*, 27(2), 375–389.
- Goodwin, T. (2012). Why we should reject “nudge.” *Politics*, 32(2), 85–92. <https://doi.org/10.1111/j.1467-9256.2012.01430.x>
- Gorton, D., Ni Mhurchu, C., Chen, M., & Dixon, R. (2009). Nutrition labels: a survey of use, understanding and preferences among ethnically diverse shoppers in New Zealand. *Public Health Nutrition*, 12(09), 1359. <https://doi.org/10.1017/S1368980008004059>
- Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, 32(3), 463–483. <https://doi.org/10.1017/S0266267115000425>
- Grüne-Yanoff, T., & Hertwig, R. (2015). Nudge Versus Boost: How Coherent are Policy and Theory? *Minds and Machines*, 26(1–2), 1–52. <https://doi.org/10.1007/s11023-015-9367-9>
- Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public Views on Policies Involving Nudges. *Review of Philosophy and Psychology*, 6(3), 439–453. <https://doi.org/10.1007/s13164-015-0263-2>
- Hahn, U. (2011). The problem of circularity in evidence, argument, and explanation. *Perspectives on Psychological Science*, 6(2), 172–182.
- Haines, M. P., Barker, G., & Rice, R. (2003). The NIU Experiment: A case study of the social norms approach. In H. W. Perkins (Ed.), *The Social Norms Approach to Preventing School and College Age Substance Abuse: A Handbook for Educators, Counselors, and Clinicians* (p. 22). San Francisco, Calif: Jossey-Bass.
- Haines, M. P., & Spear, S. F. (1996). Changing the perception of the norm: a strategy to decrease binge drinking among college students. *Journal of American College Health : J of ACH*, 45(3), 134–140. <https://doi.org/10.1080/07448481.1996.9936873>
- Hall, M. G., Marteau, T. M., Sunstein, C. R., Ribisl, K. M., Noar, S. M., Orlan, E. N., & Brewer, N. T. (2018). Public support for pictorial warnings on cigarette packs: an experimental study of US smokers. *J Behav Med*. <https://doi.org/10.1007/s10865-018-9910-2>
- Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., Ekelund, U., & Group., L. P. A. S. W. G. (2012). Global physical activity levels: surveillance progress, pitfalls, and prospects. *The Lancet*, 380(9838), 247–257.
- Hammond, D., Daniel, S., & White, C. M. (2013). The effect of cigarette branding and plain packaging on female youth in the United Kingdom. *The Journal of Adolescent Health* :

- Official Publication of the Society for Adolescent Medicine*, 52(2), 151–157.
<https://doi.org/10.1016/j.jadohealth.2012.06.003>
- Hammond, D., Fong, G. T., McNeill, A., Borland, R., Cummings, K. M., & Hammond, D. (2006). Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: findings from the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control*, 15, 19–25. <https://doi.org/10.1136/tc.2005.012294>
- Hammond, D., & Parkinson, C. (2009). The impact of cigarette package design on perceptions of risk. *Journal of Public Health*, 31(3), 345–353. <https://doi.org/10.1093/pubmed/fdp066>
- Hancock, L., Abhold, J., Gascoigne, J., & Altekruze, M. (2002). Applying Social Norms Marketing to Tobacco Cessation and Prevention: Lessons Learned from Three Campaigns. *The Report on Social Norms, Working Pa.* Retrieved from https://www.mendeley.com/research/applying-social-norms-marketing-tobacco-cessation-prevention-lessons-learned-three-campaigns/?utm_source=desktop&utm_medium=1.15&utm_campaign=open_catalog&userDocumentId=%7B49f39eea-2ba5-40fa-b274-05ed28604d4d%7D
- Hancock, L., & Henry, N. (2003). *Perceptions, Norms and Tobacco Use in College Residence Hall Freshmen: Evaluation of a Social Norms Marketing Intervention*. San Francisco: Jossery-Bass.
- Hanna, S. D., Kim, K. T., & Chen, S. C. C. (2016). Retirement savings. In J. J. Xiao (Ed.), *Handbook of Consumer Finance Research* (pp. 33–43). New York: Springer International Publishing.
- 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*, 1, 3–28.
- Harnack, L. J., & French, S. A. (2008). Effect of point-of-purchase calorie labeling on restaurant and cafeteria food choices: A review of the literature. *International Journal of Behavioral Nutrition and Physical Activity*, 5(1), 51. <https://doi.org/10.1186/1479-5868-5-51>
- Harris, A. J. L., Corner, A., & Hahn, U. (2013). James is polite and punctual (and useless): A Bayesian formalisation of faint praise. *Thinking & Reasoning*, 19(3–4), 414–429. <https://doi.org/10.1080/13546783.2013.801367>
- Hausman, D. M., & Welch, B. (2010). Debate: To Nudge or Not to Nudge. *The Journal of Political Philosophy*, 18(1), 123–136.

- Hedlin, S., & Sunstein, C. R. (2016). Does Active Choosing Promote Green Energy Use: Experimental Evidence. *Ecology*, *43*, 107.
- Hogg, M. (2011). *Do we need a deeper, more complex conversation with the public about global issues? A review of literature.*
- Holden, S. S., Zlatevska, N., & Dubelaar, C. (2016). Whether Smaller Plates Reduce Consumption Depends on Who's Serving and Who's Looking: A Meta-Analysis. *Journal of the Association for Consumer Research*, *1*(1), 134–146.
- Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Kelly, M. P., Nakamura, R., ... Ogilvie, D. (2013). Altering micro-environments to change population health behaviour: towards an evidence base for choice architecture interventions. *BMC Public Health*, *13*(1), 1218. <https://doi.org/10.1186/1471-2458-13-1218>
- Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Lewis, H. B., Wei, Y., ... Ogilvie, D. (2015). Portion, package or tableware size for changing selection and consumption of food, alcohol and tobacco. In G. J. Hollands (Ed.), *Cochrane Database of Systematic Reviews*. UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/14651858.CD011045.pub2>
- Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Kelly, M. P., Nakamura, R., ... Ogilvie, D. (2013). *Altering choice architecture to change population health behaviour: a large-scale conceptual and empirical scoping review of interventions within micro-environments*. Cambridge.
- Holmes, A. S., Serrano, E. L., Machin, J. E., Duetsch, T., & Davis, G. C. (2013). Effect of different children's menu labeling designs on family purchases. *Appetite*, *62*, 198–202. <https://doi.org/10.1016/j.appet.2012.05.029>
- Horack, S., & Wood, A. (2005). *An evaluation of scheme joining techniques in workplace pension schemes with an employer contribution: Case Study. Research Report (Vol. 292)*. London.
- House of Lords. (2011). *Behaviour Change*. London. Retrieved from <http://www.parliament.uk/business/lords>
- Hull, C. L. (1943). *Principles of behaviour: An introduction to behaviour theory*. New York: Appleton-Century Crofts.
- Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2016). When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects. *Available at SSRN*.
- Jog, V., & Lee, I. (2016). Reforming Canada's retirement savings system-solutions for a non-

- existent problem? *Public Finance & Management*, 16(2).
- John, P., Moseley, A., Wales, C., Smith, G., Richardson, L., Stoker, G., & Cotterill, S. (2013). *Nudge, Nudge, Think, Think: Experimenting with Ways to Change Civic Behaviour*. London: Bloomsbury Academic.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults, Framing and Privacy: Why Opting In-Opting Out. *Marketing Letters*, 13(1), 5–15.
- Johnson, E. J., & Goldstein, D. G. (2003). Do defaults save lives? *Science*, 302(5649), 1338–1339. <https://doi.org/10.1126/science.1091721>
- Jolls, C., Sunstein, C. R., & Thaler, R. (1998). *A Behavioral Approach to Law and Economics*. *Stanford Law Review*. Retrieved from <https://ssrn.com/abstract=74927>
- Jung, J. Y., & Mellers, B. A. M. (2016). American attitudes toward nudges. *Judgment and Decision Making*. *Judgement and Decision Making*, 11(1), 62–74.
- Junghans, A. F., Cheung, T. T. L., & De Ridder, D. T. D. (2015). Under consumers' scrutiny - an investigation into consumers' attitudes and concerns about nudging in the realm of health behavior. *BMC Public Health*, 15, 336. <https://doi.org/10.1186/s12889-015-1691-8>
- Kahneman, D. (2003). Maps of bounded rationality: a perspective on intuitive judgement and choice. In T. Frangmyr (Ed.), *Lex Prix Nobel: The Nobel Prizes 2002* (pp. 449–489). Stockholm: Almqvist & Wiksell International.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49–81). New York: Cambridge University Press.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoack & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267–293). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 493–508). New York: Cambridge University Press. [https://doi.org/10.1016/0010-0277\(82\)90022-1](https://doi.org/10.1016/0010-0277(82)90022-1)
- Kerr, J., Eves, F., & Carroll, D. (2001a). Six-Month Observational Study of Prompted Stair Climbing. *Preventive Medicine*, 33(5), 422–427. <https://doi.org/10.1006/pmed.2001.0908>
- Kerr, J., Eves, F. F., & Carroll, D. (2001b). Encouraging stair use: Stair riser banners are better

- than poster prompts. *American Journal of Public Health*, 91(8), 1192–1193. Retrieved from https://www.mendeley.com/research/encouraging-stair-stair-riser-banners-better-poster-prompts/?utm_source=desktop&utm_medium=1.15&utm_campaign=open_catalog&userDocumentId=%7Be50beb00-7aee-4cfa-a07e-eb2b0e721b86%7D
- Kerr, J., Eves, F. F., & Carroll, D. (2001c). Getting more people on the stairs: the impact of a new message format. *Journal of Health Psychology*, 6(5), 495–500. <https://doi.org/10.1177/135910530100600503>
- Kiszko, K. M., Martinez, O. D., Abrams, C., & Elbel, B. (2014). The Influence of Calorie Labeling on Food Orders and Consumption: A Review of the Literature. *Journal of Community Health*, 39(6), 1248–1269. <https://doi.org/10.1007/s10900-014-9876-0>
- Klaczynski, P. A. (2001). Analytic and Heuristic Processing Influences on Adolescent Reasoning and Decision Making. *Child Development*, 72(3), 844–861.
- Knoef, M., Been, J., Alessie, R., Caminada, K., Goudswaard, K., & Kalwij, A. (2016). Measuring retirement savings adequacy: developing a multi-pillar approach in the Netherlands. *Journal of Pension Economics and Finance*, 15(1), 55–89.
- Kosters, M., & Van der Heijden, J. (2015). From mechanism to virtue: Evaluating Nudge theory. *Evaluation*, 21(3), 276–291. <https://doi.org/10.1177/1356389015590218>
- Kruglanski, A. W. (2013). Only One? The Default Interventionist Perspective as a Unimodel-Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, 8(3), 242–247. <https://doi.org/10.1177/1745691613483477>
- Lagnado, D. A., Fenton, N., & Neil, M. (2012). Legal idioms: a framework for evidential reasoning. *Argument & Computation*, 4(1), 46–63. <https://doi.org/10.1080/19462166.2012.682656>
- Laplace, S. (1951). *A philosophical essay on probabilities (In F.W. Truscott & F. L. Emory Trans Ed.)*. New York: Dover (Original Work Published 1814).
- Laszlo, E. (1989). *The Inner Limits of Mankind*. London: One World.
- Lee, K. K., Perry, A. S., Wolf, S. A., Agarwal, R., Rosenblum, R., Fischer, S., ... Silver, L. D. (2012). Promoting Routine Stair Use Evaluating the Impact of a Stair Prompt Across Buildings. *American Journal of Preventive Medicine*, 42(2), 136–141. <https://doi.org/10.1016/j.amepre.2011.10.005>
- Lewis, A., & Eves, F. (2012). Prompt before the choice is made: Effects of a stair-climbing intervention in university buildings. *British Journal of Health Psychology*, 17(3), 631–

643. <https://doi.org/10.1111/j.2044-8287.2011.02060.x>
- Lewis, M. A., & Neighbors, C. (2006). Social norms approaches using descriptive drinking norms education: a review of the research on personalized normative feedback. *Journal of American College Health, 54*(4), 213–218. <https://doi.org/10.3200/JACH.54.4.213-218>
- Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, Effectiveness, and Ethics. *Basic and Applied Social Psychology, 39*(6), 293–306. <https://doi.org/https://doi.org/10.1080/01973533.2017.1356304>
- Lin, Y., Osman, M., Harris, A. J. L., & Read, D. (2018). Underlying Wishes and Nudged Choices. *JEP: Applied*.
- Linkenbach, J., Perkins, H. W., & DeJong, W. (2003). *Parent's Perceptions of Parenting Norms: Using the Social Norms Approach to Reinforce Effective Parenting*. San Francisco: Jossey-Bass.
- Liu, P. J., Wisdom, J., Roberto, C. A., Liu, L. J., & Ubel, P. A. (2014). Using Behavioral Economics to Design More Effective Food Policies to Address Obesity. *Applied Economic Perspectives and Policy, 36*(1), 6–24. <https://doi.org/10.1093/aep/ppt027>
- Loewenstein, G., Asch, D. A., Friedman, J. Y., Melichar, L. A., Volpp, K. G., Loewenstein, G., ... Volpp, K. G. (2012). Can behavioural economics make us healthier? *BMJ, 344*, e3482. <https://doi.org/10.1136/bmj.e3482>
- Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2014). Warning: You are About to Be Nudged. *SSRN Electronic Journal, 1*–23. <https://doi.org/10.2139/ssrn.2417383>
- Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2015). Warning: You are about to be nudged. *Behavioral Science & Policy, 1*(1), 35–42.
- Loosemore, T. (2014). One link on GOV.UK - 350,000 more organ donors. Retrieved from <https://gds.blog.gov.uk/2014/03/18/organ-donor-register/>
- Lusardi, A., & Mitchell, O. S. (2011a). Financial literacy and planning: Implications for retirement wellbeing (No. w17078). In A. Lusardi & O. S. Mitchell (Eds.), *Financial Literacy. Implications for Retirement Security and the Financial Marketplace* (pp. 17–39). New York: Oxford University Press.
- Lusardi, A., & Mitchell, O. S. (2011b). Financial literacy around the world: an overview. *Journal of Pension Economics and Finance, 10*(4), 497–508.
- Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature, 52*(1), 5–44.
- Magnusson, R. S. (2009). Rethinking global health challenges: towards a 'global compact' for

- reducing the burden of chronic disease. *Public Health*, 123(3), 265–274.
- Mandell, L., & Klein, L. S. (2007). Motivation and financial literacy. *Financial Services Review*, 16(2), 105.
- Marshall, A. L., Bauman, A. E., Patch, C., Wilson, J., & Chen, J. (2002). Can motivational signs prompt increases in incidental physical activity in an Australian health-care facility? *Health Education Research*, 17(6), 743–749. <https://doi.org/10.1093/her/17.6.743>
- Marteau, T. M., Ogilvie, D., Roland, M., Suhrcrke, M., & Kelly, M. P. (2011). Judging nudging: can nudging improve population health? *BMJ (Clinical Research Ed.)*, 342, d228. <https://doi.org/10.1136/bmj.d228>
- Maynard, O. M., Munafò, M. R., & Leonards, U. (2013). Visual attention to health warnings on plain tobacco packaging in adolescent smokers and non-smokers. *Addiction*, 108(2), 413–419. <https://doi.org/10.1111/j.1360-0443.2012.04028.x>
- Mazzocchi, M., Cagnone, S., Bech-Larsen, T., Niedźwiedzka, B., Saba, A., Shankar, B., ... Traill, W. B. (2015). What is the public appetite for healthy eating policies? Evidence from a cross-European survey. *Health Economics, Policy and Law*, 10(03), 267–292. <https://doi.org/10.1017/S1744133114000346>
- McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, 17(5), 414–420. <https://doi.org/10.1111/j.1467-9280.2006.01721.x>
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(42), 1–11.
- Mill, J. S. (1863). *On Liberty*. Boston, MA: Ticknor and Fields.
- Mills, C. (2013). Why nudges matter: a reply to Goodwin. *Politics*, 33(1), 28–36. <https://doi.org/10.1111/j.1467-9256.2012.01450.x>
- 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), 195. <https://doi.org/10.1111/1475-6765.12073>
- Moodie, C., & Mackintosh, A. M. (2013). Young adult women smokers' response to using plain cigarette packaging: a naturalistic approach. *BMJ Open*, 3(3), e002402. <https://doi.org/10.1136/bmjopen-2012-002402>
- Moodie, C., Mackintosh, A. M., Hastings, G., & Ford, A. (2011). Young adult smokers' perceptions of plain packaging: a pilot naturalistic study. *Tobacco Control*, 20(5), 367–

373. <https://doi.org/10.1136/tc.2011.042911>
- Moodie, C., Stead, M., Bauld, L., McNeill, A., Angus, K., Hinds, K., ... O'Mara-Eves, A. (2012). *Plain Tobacco Packaging: A Systematic Review*. London. Retrieved from http://dspace.stir.ac.uk/bitstream/1893/24420/1/Moodie_et_al_2012_PHRC_PlainPacks.pdf
- Mossialos, E., Costa-Font, J., & Rudisill, C. (2008). Does organ donation legislation affect individuals' willingness to donate their own or their relative's organs? Evidence from European Union survey data. *BMC Health Services Research*, 8(1), 48. <https://doi.org/10.1186/1472-6963-8-48>
- Munafò, M. R., Roberts, N., Bauld, L., & Leonards, U. (2011). Plain packaging increases visual attention to health warnings on cigarette packs in non-smokers and weekly smokers but not daily smokers. *Addiction*, 106(8), 1505–1510. <https://doi.org/10.1111/j.1360-0443.2011.03430.x>
- Mutran, E. J., Reitzes, D. C., & Fernandez, M. E. (1997). Factors that influence attitudes toward retirement. *Research on Aging*, 19(3), 251–273.
- National Advisory Committee on Bioethics. (2015). *Nudging in Public Health – An Ethical Framework*.
- NHSBT. (2015). News Release: Call for everyone in the UK to discuss organ donation as the number of deceased organ donors in the UK drops for the first time in more than a decade. Retrieved December 4, 2015, from http://www.nhsbt.nhs.uk/news-and-media/news-articles/news_2015_07_20.asp
- NHSBT. (2016). Highest number of organ transplants ever across UK, but many families still say no to donation. Retrieved from <https://www.organdonation.nhs.uk/news-and-campaigns/news/highest-number-of-organ-transplants-ever-across-uk-but-many-families-still-say-no-to-donation/>
- NHSBT. (2017). Organ Donation and Transplantation: Activity Report 2016/17. Retrieved from https://nhsbt.dbe.blob.core.windows.net/umbraco-assets-corp/4657/activity_report_2016_17.pdf
- Nikolaou, C. K., Hankey, C. R., & Lean, M. E. J. (2014). Preventing weight gain with calorie-labelling. *Obesity*, 22(11), 2277–2283.
- Nishida, Y., Akaoka, I., & Nishizawa, T. (1975). Effect of sex hormones on uric acid metabolism in rats. *Experientia*, 31(10), 1134–1135. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1204720>

- Nocon, M., Müller-Riemenschneider, F., Nitzschke, K., & Willich, S. N. (2010). Review Article: Increasing physical activity with point-of-choice prompts--a systematic review. *Scandinavian Journal of Public Health*, 38(6), 633–638. <https://doi.org/10.1177/1403494810375865>
- Nomura, T., Yoshimoto, Y., Akezaki, Y., & Sato, A. (2009). Changing behavioral patterns to promote physical activity with motivational signs. *Environmental Health and Preventive Medicine*, 14(1), 20–25. <https://doi.org/10.1007/s12199-008-0053-x>
- Noone, J. H. (2010). *Psychological and socioeconomic factors influencing men and women's planning for retirement: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Psychology at Massey University, Turitea, New Zealand (Doct.*
- Nørnberg, T. R., Houlby, L., Skov, L., & Pérez-Cueto, F. J. (2016). Choice architecture interventions for increased vegetable intake and behaviour change in a school setting: a systematic review. *Perspect Public Health*, 136(3), 134–142. <https://doi.org/10.1177/1757913915596017>
- OECD. (2013). *Pensions at a glance 2013: OECD and G20 indicators*. [https://doi.org/OECD\(2013\), Pensions at a Glance 2013: OECD and G20 Indicators](https://doi.org/OECD(2013),Pensions%20at%20a%20Glance%202013%3A%20OECD%20and%20G20%20Indicators), OECD Publishing. http://dx.doi.org/10.1787/pension_glance-2013-en
- Office of Fair Trading. (2013). *Defined contribution workplace pension market study*. Retrieved from http://webarchive.nationalarchives.gov.uk/20131101164215/http://www.oft.gov.uk/shared_of/market-studies/oft1505;jsessionid=BA96A643DCEA2A81D543CF38CB586719
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, 12(6), 237–241. <https://doi.org/10.1016/j.tics.2008.02.014>
- Organdonor.gov. (2017). The Need is Real: Data. Retrieved February 20, 2016, from <http://www.organdonor.gov/about/data.html>
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11(6), 988–1010. <https://doi.org/10.3758/BF03196730>
- Osman, M. (2007). Can tutoring improve performance on a reasoning task under deadline conditions? *Memory & Cognition*, 35(2), 342–351.
- Osman, M. (2013). A Case Study: Dual-Process Theories of Higher Cognition-Commentary on Evans & Stanovich (2013). *Prospect Psychological Science*, 8(3), 248–252. <https://doi.org/10.1177/1745691613483475>

- Osman, M. (2014a). *Future-minded: The Psychology of Agency and Control*. New York: Palgrave Macmillan.
- Osman, M. (2014b). What are the essential cognitive requirements for prospection (thinking about the future)? *Frontiers in Psychology*, 5, 626. <https://doi.org/10.3389/fpsyg.2014.00626>
- Osman, M. (2016). Nudge: How Far Have We Come? *Æconomia. History, Methodology, Philosophy*, (6–4), 557–570. Retrieved from <https://oeconomia.revues.org/2490>
- Osman, M., & Stavy, R. (2006). Intuitive rules: From formative to developed reasoning. *Psychonomic Bulletin & Review*, 13(935–53).
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.
- Pearson, M., & West, P. (2003). Drifting smoke rings: Social network analysis and Markov processes in a longitudinal study of friendship groups and risk-taking. *Connections*, 25(2), 59–76.
- Pechey, R., Burge, P., Mentzakis, E., Suchrcke, M., & Marteau, T. M. (2014). Public acceptability of population-level interventions to reduce alcohol consumption: A discrete choice experiment. *Social Science & Medicine*, 113, 104–109.
- Perkins, H. W., Meilman, P. W., Leichter, J. S., Cashin, J. R., & Presley, C. A. (1999). Misperceptions of the Norms for the Frequency of Alcohol and Other Drug Use on College Campuses. *Journal of American College Health*, 47(6), 253–258. <https://doi.org/10.1080/07448489909595656>
- 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.
- Puig-Ribera, A., & Eves, F. F. (2010). Promoting stair climbing in Barcelona: similarities and differences with interventions in English-speaking populations. *European Journal of Public Health*, 20(1), 100–102. <https://doi.org/10.1093/eurpub/ckp059>
- Radecki, C. M., & Jaccard, J. (1997). Psychological aspects of organ donation: a critical review and synthesis of individual and next-of-kin donation decisions. *Health Psychology*, 16(2), 183–195.
- Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, parallel constraint satisfaction processes, and gestalt principles: (Re)introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review*, 1, 26–53.

- Rebonato, R. (2013). A Critical Assessment of Libertarian Paternalism.
- Reisch, L. A., & Sunstein, C. R. (2016). Do Europeans like nudges? *Judgement and Decision Making*, *11*(4), 310–325.
- Reisch, L. A., Sunstein, C. R., & Gwozdz, W. (2016). Better Than A Whip? European Attitudes Toward Health Nudges. *Food Policy*, *69*, 1–10. <https://doi.org/doi.org/10.1016/j.foodpol.2017.01.007>
- Rithalia, A., McDaid, C., Suekarran, S., Norman, G., Myers, L., & Sowden, A. (2009). A systematic review of presumed consent systems for deceased organ donation. *Health Technology Assessment*, *13*(26). <https://doi.org/10.3310/hta13260>
- Robinson, E., Nolan, S., Tudur-Smith, C., Boyland, E. J., Harrold, J. A., Hardman, C. A., & Halford, J. C. G. (2014). Will smaller plates lead to smaller waists? A systematic review and meta-analysis of the effect that experimental manipulation of dishware size has on energy consumption. *Obesity Reviews*, *15*(10), 812–821. <https://doi.org/10.1111/obr.12200>
- Rosenblum, A. M., Horvat, L. D., Siminoff, L. A., Prakash, V., Beitel, J., & Garg, A. X. (2012). The authority of next-of-kin in explicit and presumed consent systems for deceased organ donation: an analysis of 54 nations. *Nephrology Dialysis Transplantation*, *27*(6), 2533–2546. <https://doi.org/10.1093/ndt/gfr619>
- Russell, W. D., Dzewaltowski, D. A., & Ryan, G. J. (1999). The Effectiveness of a Point-of-Decision Prompt in Deterring Sedentary Behavior. *American Journal of Health Promotion*, *13*(5), 257–259. <https://doi.org/10.4278/0890-1171-13.5.257>
- Sacks, G., Rayner, M., & Swinburn, B. (2009). Impact of front-of-pack “traffic-light” nutrition labelling on consumer food purchases in the UK. *Health Promotion International*, *24*(4), 344–352. <https://doi.org/10.1093/heapro/dap032>
- Saghai, Y. (2013). Salvaging the concept of nudge. *Journal of Medical Ethics*, *39*(8), 487–493. <https://doi.org/10.1136/medethics-2012-100727>
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*(3), 414.
- Schwarz, N. (1998). Warmer and More Social: Recent Developments in Cognitive Social Psychology. *Annual Review of Sociology*, *24*(1), 239–264. <https://doi.org/10.1146/annurev.soc.24.1.239>
- Science and Technology Committee. (2014). *Letter to the House of Lords*. London. Retrieved

- from <http://www.parliament.uk/documents/lords-committees/science-technology/behaviourchange/LetwinBehaviourChangeLtr20140722.pdf>
- Selinger, E., & White, K. P. (2012). What Counts as a Nudge? *The American Journal of Bioethics*, *12*(2), 11–12. <https://doi.org/10.1080/15265161.2011.634485>
- Selinger, E., & Whyte, K. P. (2012). Nudging cannot solve complex policy problems. *European Journal of Risk Regulation*, *3*(1), 26–31.
- Shepherd, L., O’Carroll, R. E., & Ferguson, E. (2014). An international comparison of deceased and living organ donation/transplant rates in opt-in and opt-out systems: a panel study. *BMC Medicine*, *12*(1), 131. <https://doi.org/10.1186/s12916-014-0131-4>
- Shu, S. B., Bang, M. H., & Weber, E. U. (2016). Informed Consent to Choice Architecture: Transparency, Trust, and Perceived Effectiveness. *In Preparation*.
- Shu, S. B., & Payne, J. (2015). Psychological Factors in Savings and Decumulation Decisions. *NA-Advances in Consumer Research*, *43*.
- Siminoff, L. A., & Lawrence, R. H. (2002). Knowing Patients’ Preferences about Organ Donation: Does it Make a Difference? *The Journal of Trauma*, *53*(4), 754–760. <https://doi.org/10.1097/01.TA.0000030052.69240.4F>
- Simon, D., & Holyoak, K. J. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review*, *6*, 283–294.
- Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814–837.
- Sinclair, S. E., Cooper, M., & Mansfield, E. D. (2014). The Influence of Menu Labeling on Calories Selected or Consumed: A Systematic Review and Meta-Analysis. *Journal of the Academy of Nutrition and Dietetics*, *114*(9), 1375–1388. <https://doi.org/10.1016/j.jand.2014.05.014>
- Skov, L. R., Lourenço, S., Hansen, G. L., Mikkelsen, B. E., & Schofield, C. (2013). Choice architecture as a means to change eating behaviour in self-service settings: a systematic review. *Obesity Reviews*, *14*(3), 187–196. <https://doi.org/10.1111/j.1467-789X.2012.01054.x>
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22. <https://doi.org/10.1037/0033-2909.119.1.3>
- Soler, R. E., Leeks, K. D., Buchanan, L. R., Brownson, R. C., Heath, G. W., Hopkins, D. H., & the Task Force on Community Preventive Services. (2010). Point-of-Decision Prompts

- to Increase Stair Use A Systematic Review Update. *American Journal of Preventive Medicine*, 38(2 Suppl), S292–S300. <https://doi.org/10.1016/j.amepre.2009.10.028>
- Song, C. (2015). Financial illiteracy and pension contributions: A field experiment on compound interest in China. Retrieved from <http://dx.doi.org/10.2139/ssrn.2580856>.
- Sonnenberg, L., Gelsomin, E., Levy, D. E., Riis, J., Barraclough, S., & Thorndike, A. N. (2013). A traffic light food labeling intervention increases consumer awareness of health and healthy choices at the point-of-purchase. *Preventive Medicine*, 57(4), 253–257. <https://doi.org/10.1016/j.ypmed.2013.07.001>
- Spital, A. (1993). A consent for organ donation: Time for change. *Clinical Transplant*, 7, 525–528.
- Spital, A. (1995). Mandated choice: A plan to increase public commitment to organ donation. *JAMA*, 273(6), 504–506. <https://doi.org/10.1001/jama.1995.03520300078045>
- Spital, A. (1996). Mandated choice for organ donation: Time to give it a try. *Annals of Internal Medicine*, 125(1), 66–69.
- Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, 23(5), 701–717. <https://doi.org/10.1017/S0140525X00623439>
- Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychological Review*, 112(1), 159.
- Sunstein, C. R. (1993). Endogenous Preferences, Environmental Law. *The Journal of Legal Studies*, 22(2), 217–254.
- Sunstein, C. R. (2013). *Simple: The Future of Government*. New York: Simon & Schuster.
- Sunstein, C. R. (2014a). Nudging: A Very Short Guide. *Journal of Consumer Policy*, 37, 1–8.
- Sunstein, C. R. (2014b). *Why Nudge? The Politics of Libertarian Paternalism*. New Haven: Yale University Press.
- Sunstein, C. R. (2015a). *Choosing not to choose: Understanding the value of choice*. New York, NY: Oxford University Press.
- Sunstein, C. R. (2015b). Nudges, Agency, and Abstraction: A Reply to Critics. *Review of Philosophy and Psychology*, 6(3), 511–529. <https://doi.org/10.1007/s13164-015-0266-z>
- Sunstein, C. R. (2015c). The Ethics of Nudging. *Yale Journal on Regulation*, 32(2), 413–450.
- Sunstein, C. R. (2016a). Autonomy by Default. *The American Journal of Bioethics*, 16(11), 1–2. <https://doi.org/https://doi.org/10.1080/15265161.2016.1222169>
- Sunstein, C. R. (2016b). Do People Like Nudges? *Administrative Law Review*, Forthcoming. <https://doi.org/10.2139/ssrn.2604084>

- Sunstein, C. R. (2016c). People prefer system 2 nudges (kind of). *Duke Law Journal*, 66, 121–168.
- Sunstein, C. R. (2016d). *The Ethics of Influence: Government in the Age of Behavioral Science* (ed.). Cambridge University Press.
- Sunstein, C. R., Reisch, L. A., & Rauber, J. (2017). *A World-Wide Consensus on Nudging? Not Quite, But Almost. Regulation and Governance*. <https://doi.org/10.2139/ssrn.2955693>
- Sunstein, C. R., Reisch, L. A., & Reauber, J. (2017). A worldwide consensus on nudging? Not quite, but almost. *Regulation & Governance*, 12, 3–22.
- Sunstein, C. R., & Thaler, R. H. (2003). Liberatorian paternalism is not an oxymoron. *The University of Chicago Law Review*, 70(4), 1159–1202.
- Swartz, J. J., Braxton, D., & Viera, A. J. (2011). Calorie menu labeling on quick-service restaurant menus: an updated systematic review of the literature. *International Journal of Behavioural Nutrition and Physical Activity*, 8, 135. <https://doi.org/10.1186/1479-5868-8-135>
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian Paternalism. *American Economic Review*, 93(2), 175–179.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.
- The Guardian. (2015). Organ donation rates held back by families refusing consent, study finds. Retrieved from <https://www.theguardian.com/australia-news/2015/mar/01/organ-donation-rates-held-back-by-families-refusing-consent-study-finds>
- Thombs, D. L., Dotterer, S., Olds, R. S., Sharp, K. E., & Raub, C. G. (2004). A close look at why one social norms campaign did not reduce student drinking. *Journal of American College Health*, 53(2), 61–68. <https://doi.org/10.3200/JACH.53.2.61-70>
- Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–196). Oxford, England: Oxford University Press.
- Tomer, J. F. (2017). *Advanced Introduction to Behavioral Economics*. Glos: UK: Edward Elgar Publishing Limited.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science (New Series)*, 185(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039–1061.

- U.S. Department of Health and Human Services. (2013). *2012 National Survey of Organ Donation Attitudes and Behaviors*. Rockville, Maryland. Retrieved from <http://organdonor.gov/dtcp/nationalsurveyorgandonation.pdf>
- Vallgård, S. (2012). Nudge a new and better way to improve health? *Health Psychology, 104*(2), 200–203. <https://doi.org/10.1016/j.healthpol.2011.10.013>
- van Dalen, H. P., & Henkens, K. (2014). Comparing the effects of defaults in organ donation systems. *Social Science & Medicine, 106*, 137–142. <https://doi.org/10.1016/j.socscimed.2014.01.052>
- Van Rooij, M. C., Lusardi, A., & Alessie, R. J. (2011). Financial literacy and retirement planning in the Netherlands. *Journal of Economic Psychology, 32*(4), 593–608.
- Vincent, A., & Logan, L. (2012). Consent for organ donation. *British Journal of Anaesthesia, 108*(suppl 1), i80–i87. <https://doi.org/10.1093/bja/aer353>
- Wakefield, M., Coomber, K., Zacher, M., Durkin, S., Brennan, E., & Scollo, M. (2015). Australian adult smokers' responses to plain packaging with larger graphic health warnings 1 year after implementation: results from a national cross-sectional tracking survey. *Tobacco Control, 24*(Suppl 2), ii17–ii25. <https://doi.org/10.1136/tobaccocontrol-2014-052050>
- Wansink, B. (2004). Environmental factors that increase the food intake and consumption volume of unknowing consumers. *Annual Review of Nutrition, 24*, 455–479. <https://doi.org/10.1146/annurev.nutr.24.012003.132140>
- Wansink, B., & van Ittersum, K. (2005). Shape of glass and amount of alcohol poured: comparative study of effect of practice and concentration. *BMJ, 331*. <https://doi.org/10.1136/bmj.331.7531.1512>
- Wansink, B., & van Ittersum, K. (2013). Portion size me: Plate-size induced consumption norms and win-win solutions for reducing food intake and waste. *Journal of Experimental Psychology: Applied, 19*(4), 320–332. <https://doi.org/10.1037/a0035053>
- Webb, K. L., Solomon, L. S., Sanders, J., Akiyama, C., & Crawford, P. B. (2011). Menu labeling responsive to consumer concerns and shows promise for changing patron purchases. *Journal of Hunger & Environmental Nutrition, 6*(2), 166–178. <https://doi.org/10.1080/19320248.2011.576210>
- Webb, O. J., & Eves, F. F. (2007). Promoting Stair Climbing: Intervention Effects Generalize to a Subsequent Stair Ascent. *American Journal of Health Promotion, 22*(2), 114–119. <https://doi.org/10.4278/0890-1171-22.2.114>

- Wells, J. (2014). *Pension Annuities: A review of consumer behaviour*. UK.
- White, M. (2013). *The Manipulation of Choice: Ethics and Libertairan Paternalism*. New York: Palgrave Macmillan.
- Whitman, D. G., & Rizzo, M. J. (2015). The Problematic Welfare Standards of Behavioural Paternalism. *Review of Philosophy and Psychology*, 6(3).
- WHO. (2015). Health Warnings. Retrieved March 3, 2017, from <http://www.euro.who.int/en/health-topics/disease-prevention/tobacco/policy/who-framework-convention-on-tobacco-control-who-fctc/key-areas-of-tobacco-control-policy/health-warnings>
- WHO. (2016). WHO | Tobacco. Retrieved March 15, 2017, from <http://www.who.int/mediacentre/factsheets/fs339/en/>
- WHO. (2017a). NCDs | Noncommunicable diseases and their risk factors. Retrieved March 2, 2017, from <http://www.who.int/ncds/introduction/en/>
- WHO. (2017b). *Plain packaging of tobacco products: evidence, design and implementation*. Retrieved from http://apps.who.int/iris/bitstream/10665/207478/1/9789241565226_eng.pdf?ua=1
- Whyte, K. P., Selinger, E., Caplan, A. L., & Sadowski, J. (2012). Nudge, Nudge or Shove, Shove—The Right Way for Nudges to Increase the Supply of Donated Cadaver Organs. *The American Journal of Bioethics*, 12(2), 32–39. <https://doi.org/10.1080/15265161.2011.634484>
- Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies*, 61, 341–355. <https://doi.org/10.1111/j.1467-9248.2012.00974.x>
- Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin*, 106(2), 231.
- Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in Everyday Life: Thought, Emotion, and Action. *Journal of Personality and Social Psychology*, 83(6), 1281–1297. <https://doi.org/10.1037//0022-3514.83.6.1281>
- World Bank. (2016). Pensions. Retrieved from <http://www.worldbank.org/en/topic/pensions>

APPENDIX A

CHAPTER 2: Empirical Evidence Review

Unhealthy Diet		Effectiveness
Type 1	<p><u>Wansink and van Ittersum (2013)</u></p> <ul style="list-style-type: none"> - Estimate the visual fill-level size of one’s consumption norm relative to perceive level based on 219 students visually anchor around the 70% fill level for dinnerware - The influence of consumption norm on serving and consumption in a natural eating environment based on 43 diners, selected the larger plats served themselves 52% more total food than those who selected smaller plates - The influence of large plate bias on conference buffet service behaviour involved 237 professionals were consistent with expectations, that those who were given large plates served 90% more volume of food. <p><u>Robinson et al. (2014)</u></p> <ul style="list-style-type: none"> - A meta-analysis of nine studies showed that the difference in food consumption between small and large dishware conditions was marginally statistically significant and the magnitude of effect was small, with a large amount of heterogeneity. The authors concluded that the use of smaller vs. larger dishware to limit energy consumption was premature. <p><u>Hollands et al. (2015)</u></p> <ul style="list-style-type: none"> - A meta-analysis of 56 studies (6603 participants) found a small to moderate effect of portion, package, individual unit or tableware size on increasing quantities of food consumed. This effect, if sustained, could reduce average daily energy intake by 12% to 16%. <p><u>Holden, Zlatevska, and Dubelaar (2016)</u></p> <ul style="list-style-type: none"> - A meta-analysis of 56 studies showed that varying the size of the container holding food (e.g., plate or bowl) has a substantial effect on amount self-served and/or consumed; and doubling of plate size increased the amount of self-served and/or consumed by 41%. Overall, plate size had a strong effect when participants were unaware that they were participating in a food study. 	<p>Type 1 nudges designed to change plates-size have been proven effective in reducing over-consumption in food study or natural setting such as restaurants but it is unclear how this can be sustained in a more generalizable way (i.e. positive behavioral change extending beyond the immediate nudged context).</p>

<p>Type 2</p>	<p><u>Harnack and French (2008)</u></p> <ul style="list-style-type: none"> - A review of six studies provided support for the positive influence on calorie information on food choices in a cafeteria or restaurant setting, though the magnitude of the effects seen tended to be small. <p><u>Elbel, Gyamfi, and Kersh (2011)</u></p> <ul style="list-style-type: none"> - A study of 349 children and adolescents found no evidence that labelling influenced adolescent food choices or parental food choices for children in this population. <p><u>Swartz, Braxton, and Viera (2011)</u></p> <ul style="list-style-type: none"> - Only two of the seven studies reported a statistically significant reduction in calories purchased among consumers using calorie-labelled menus. The current evidence suggests that calorie labelling does not have the intended effect of decreasing calorie purchasing or consumption. <p><u>Nikolaou, Hankey, and Lean (2014)</u></p> <ul style="list-style-type: none"> - A systematic review of seven studies on the effect on calorie-labelling on calories purchased found no overall effect, but a reduction of -124.5 kcal among those who noticed the calorie-labelling. <p><u>Gorton, Ni Mhurchu, Chen, and Dixon (2009)</u></p> <ul style="list-style-type: none"> - A study of 1525 shoppers in New Zealand found that Simple Traffic Light format led to increased ability to correctly determine if a food was healthy. <p><u>Sacks, Rayner, and Swinburn (2009)</u></p> <ul style="list-style-type: none"> - The introduction of traffic light labels increased the sales of ready-meals in the 4-weeks after intervention, though there was no association between changes in product sales and the healthiness of the product. <p><u>Kiszko, Martinez, Abrams, and Elbel (2014)</u></p> <ul style="list-style-type: none"> - A review of 31 studies assess the evidence on the effectiveness of calorie labelling at the point of purchase, of which 12 are natural experiments. The results of these studies suggest future research are required to establish effectiveness. 	<p>Type 2 nudges that use calorie labelling do not seem to significantly change consumption of healthy food. The use traffic light labelling on the other hand have been effective at increasing the consumption of healthier food items in supermarket purchases. But future researches are required to establish its long-term effectiveness.</p>
----------------------	--	---

	<p><u>Emrich et al. (2017)</u></p> <ul style="list-style-type: none"> - The calorie intake consumed by 19,915 Canadian adults were reduced by 5% after the introduction of traffic light labelling. <p><u>House of Lords (2011)</u></p> <ul style="list-style-type: none"> - A field study at Sainsbury found that sales of food items with mostly green traffic light grew 46.1% whereas those with red traffic light decreased by 24%. 	
Physical Inactivity		Effectiveness
Type 1	<p><u>Åvitsland, Solbraa, and Riiser (2017)</u></p> <ul style="list-style-type: none"> - Intervention site stair climbing at baseline (79.0%) was significantly reduced with footprints. 	There are limited evidence to establish effectiveness of Type 1 nudges employing footprints to increase physical activity.
Type 2	<p><u>Andersen, Franckowiak, Snyder, Bartlett, and Fontaine (1998)</u></p> <ul style="list-style-type: none"> - A study of the effectiveness of signs to encourage use of stairs instead of escalators using 17901 shoppers found stair use increased from 4.8% to 6.9%. <p><u>Blamey, Mutrie, and Aitchison (1995)</u></p> <ul style="list-style-type: none"> - A total of 22,275 observations at underground station found that stair use increase to the order of 15-17% during the three weeks when the intervention signs were present. <p><u>Brownell, Stunkard, and Albaum (1980)</u></p> <ul style="list-style-type: none"> - A total of 45,694 observations at a shopping mall, train station and bus terminal found that stair use more than doubled for both obese and nonobese persons during two-week period when colourful sign encouraging use of 	Type 2 nudges employing motivational intervention signs have seen small increased in stair use in the short-term and no evidence to suggest that the effect is sustained in the long-term, particularly taking into account that they are

	<p>the stairs was positioned at the stairs/escalator choice point. Stair use remained elevated for 15 consecutive days while the sign was present and decreased during a 1-month follow-up period, returned to baseline by 3 months.</p> <p><u>Kerr, Eves, and Carroll (2001)</u></p> <ul style="list-style-type: none"> - A total of 45,361 observations at a shopping mall found stair use increased significantly during intervention period and when the banners were removed, remained higher than at baseline. <p><u>Lewis and Eves (2012)</u></p> <ul style="list-style-type: none"> - A total of 14,138 observations at four university buildings found that stair climbing increased significantly when volitional components was added. <p><u>Marshall, Bauman, Patch, Wilson, and Chen (2002)</u></p> <ul style="list-style-type: none"> - Self-reported data on the use of stairs in the hospital were obtained from 53 staff increased after the first intervention, but after the invention was removed stair use decreased back towards baseline levels. <p><u>Nomura, Yoshimoto, Akezaki, and Sato (2009)</u></p> <ul style="list-style-type: none"> - A total of 43,241 observations on the effectiveness of motivational signs in promoting stair use instead of escalators found that stair use increased significantly from 3.58 to 4.93% during the intervention period of 1-2 weeks, 5.80% during 2-4 weeks. <p><u>Webb and Eves (2007)</u></p> <ul style="list-style-type: none"> - A 13-week intervention in which banners carry health promotion message were introduced at intervention site found stair climbing increased by 161% and remained significantly elevated 5 weeks after the banner was removed. 	<p>unlikely to lead to any meaningful health benefits.</p>
Harmful Use of Alcohol		Effectiveness
Type 1	<p><u>Wansink and van Ittersum (2005)</u></p> <ul style="list-style-type: none"> - To determine whether people pour different amounts into short, wide glasses and into tall, slender ones. 198 college students and 86 bartenders poured more into short, wide glasses than tall slender glasses. 	<p>The effectiveness of Type 1 nudges is unclear,</p>

	<p><u>Attwood, Scott-Samuel, Stothart, Munafò, and Campanella (2012)</u></p> <ul style="list-style-type: none"> - The influence of glass shape on the rate of consumption of alcoholic and non-alcoholic beverages of 159 customers found that they were 60% slower to consume an alcoholic beverage from a straight glass compared to a curved glass. This effect was not observed for non-alcoholic beverages. <p><u>Hollands et al. (2015)</u></p> <ul style="list-style-type: none"> - One study found that adults provided with shorter, wider bottles drank larger amounts of water from them, having already poured more, compared with those provided with taller, narrower bottles. However, these was not enough evidence to estimate the effect of shape of glassware on consumption. 	<p>especially with alcoholic beverages.</p>
<p>Type 2</p>	<p><u>Bullers, Cooper, and Russell (2001)</u></p> <ul style="list-style-type: none"> - A study of the association between individuals' drinking patterns and the drinking patterns of their social network members on a sample of 1933 adults suggest that selection and influence affect the association between individual and network drinking patterns. <p><u>Perkins, Meilman, Leichliter, Cashin, and Presley (1999)</u></p> <ul style="list-style-type: none"> - Data from surveys of students representing 100 college students found that they typically misperceived their peer norms by substantially overestimating how often the average student used drugs (such as alcohol, tobacco, marijuana, etc). <p><u>Gomberg, Schneider, and DeJong (2001)</u></p> <ul style="list-style-type: none"> - A social marketing campaign to change the perception of peer drinking norms among University students found that exposure to be campaign may be associated with lower estimates of student drinking norms. <p><u>Haines and Spear (1996)</u></p> <ul style="list-style-type: none"> - A 5-year study of 23,000 student using an intervention to change perceptions of drinking norms found that the amount of binge drinking showed an 18.5% drop in the number of students who perceived binge drinking as the norm and a corresponding reduction in self-reported binge drinking of 8.8%. the apparent effectiveness of this 	<p>The evidence for the effectiveness of Type 2 nudges are encouraging at establishing the association between drinking pattern and peer drinking norms, with the effect of social norm sustained beyond the immediate short-term of 4+ months post-intervention.</p>

	<p>prevention effort suggested that changing college students' perceptions of drinking norms may lower the proportion of students who engage in binge drinking.</p> <p><u>Thombs, Dotterer, Olds, Sharp, and Raub (2004)</u></p> <ul style="list-style-type: none"> - A study of social norm campaign based on 616 students before the campaign's implementation and a follow-up survey of 723 students 4 academic years later, found that 66.5% of the students were aware of the campaign yet survey revealed no reduction in perceived drinking norm or alcohol use in this group. Further analysis revealed that (1) a majority of the students did not find the statistics used in the campaign message credible, (2) higher levels of alcohol use predicted lower levels of perceived campaign credibility, and (3) only 38.5% of the students understood the campaign's intended purpose. <p><u>Foxcraft, Moreira, Santimano, and Smith (2015)</u></p> <ul style="list-style-type: none"> - A total of 66 studies (43,1235 participants) included in the review, and 59 studies (40,951 participants) in the meta-analyses to determine whether social norm intervention reduce alcohol-related negative consequences, alcohol misuse or alcohol consumption found that the effect on social norm was sustained beyond the immediate short-term at 4+ months post intervention. 	
Tobacco Use		Effectiveness
Type 1	<p><u>Hollands et al. (2015)</u></p> <ul style="list-style-type: none"> - Three studies found no effect of longer compared with shorter cigarettes on the amounts of tobacco consumed. 	<p>There are limited evidence to establish the effectiveness of shorter cigarettes at reducing tobacco use.</p>

<p>Type 2</p>	<p><u>Hammond et al. (2006)</u></p> <ul style="list-style-type: none"> - A telephone survey conducted with 9058 adult smokers from USA, UK, Canada and Australia found that smokers in the four countries exhibited significant gaps in their knowledge of the risks of smoking. Smokers who noticed the warnings were significantly more likely to endorse health risks, including lung cancer and heart disease. Overall, smokers are not fully informed about the risks of smoking. Warnings that are graphic, larger, and more comprehensive in content are more effective in communicating the health risks of smoking. <p><u>Hammond, Daniel, and White (2013)</u></p> <ul style="list-style-type: none"> - An online survey with 947 participants rated different packs on measures of appeal and health risk, positive smoker image, and completed a behaviour pack selection task. Plain packs were rated as the least appealing and worst tasting compared with all other conditions. Respondents are also significantly less likely to accept a pack of cigarettes when offered only plain versus branded packs. <p><u>Moodie et al. (2012)</u></p> <ul style="list-style-type: none"> - A 4-week study showed that 48 smokers rated plain packaging with increased negative perceptions and feelings about the pack and about smoking as compared with branded packaging. Plain packaging also increased avoidant behaviour, certain smoking cessation behaviours, such as smoking less around others and forgoing cigarettes, and thinking about quitting. <p><u>Moodie and Mackintosh (2013)</u></p> <ul style="list-style-type: none"> - A survey study on 187 young women smokers found that in comparison to fully branded packaging, plain packaging associated with more negative perceptions and feelings about the pack and about smoking. No significant overall difference in salience, seriousness or believability of health warnings were found between the pack types, but participants reported looking more closely at the warnings on plain packs and also thinking about what the warnings were telling them as well as being more likely to take on cessation behaviour. <p><u>Brose, Chong, Aspinall, Michie, and McEwen (2014)</u></p> <ul style="list-style-type: none"> - Following abstinence of at least 12 h, 98 regular and occasional smokers were randomised to exposure to their own cigarette package, another branded package or a standardised package found that there was no significant main effect of or interaction for motivation to stop smoking. The standardised pack was perceived to be 	<p>The evidence for Type 2 nudges involving graphic health warnings seem promising, especially with the introduction of plain packaging to enhance the visibility of health warnings.</p>
----------------------	---	---

	<p>significantly less appealing and less motivating to buy cigarettes, smokers using them were perceived as less popular and cigarettes from them expected to taste worse.</p> <p><u>Dunlop, Dobbins, Young, Perez, and Currow (2014)</u></p> <ul style="list-style-type: none"> - A survey with 15,745 adults had a significant increase in cognitive and avoidant responses to on-pack health warnings after the introduction of new plain packaging. Similarly, there was a significant increase in the proportion of smokers strongly disagreeing that the look of the cigarette pack is attractive, says something good about them, influences the brand they buy, makes their pack stand out, is fashionable and matches their style. Changes in these outcomes were maintained 6 months postintervention. <p><u>Wakefield et al. (2015)</u></p> <ul style="list-style-type: none"> - A survey with 2,716 cigarette smokers one year after the introduction of plain packaging found that more smokers dislike their pack, perceived lower pack appeal, lower cigarette quality, lower satisfaction and lower value and disagree brands differed in prestige. More smokers noticed graphic health warnings, attributed motivation to quit due to these warnings, avoided specific graphic health warnings when purchasing and covered packs, with no change in perceived exaggerations of harms. 	
--	--	--

APPENDIX A: References

- Abadie, A., & Gay, S. (2006). The impact of presumed consent legislation on cadaveric organ donation: a cross-country study. *Journal of Health Economics*, 25(4), 599–620. <https://doi.org/10.1016/j.jhealeco.2006.01.003>
- Ameriks, J., Caplin, A., & Leahy, J. (2003). Wealth accumulation and the propensity to plan. *Quarterly Journal of Economics*, 118(3), 1007–1047.
- Andersen, R. E., Franckowiak, S. C., Snyder, J., Bartlett, S. J., & Fontaine, K. R. (1998). Can Inexpensive Signs Encourage the Use of Stairs? Results From a Community Intervention. *Annals of Internal Medicine*, 129(5), 363–369.
- Anthes, W. L. (2004). Financial Illiteracy in America: A Perfect Storm, a Perfect Opportunity. *Journal of Financial Service Professionals*, 58(6), 49–56.
- Appelt, K. C., Gao, J., Johnson, E. J., & von Glahn, T. (2014). Choosing How to Choose: Can People Choose the Best Choice Architecture? *Association for Consumer Research*, 42, 295–396.
- Arad, A., & Rubinstein, A. (2015). *The People's Perspective on Libertarian-Paternalistic Policies*.
- Ashcroft, R. E. (2013). Doing good by stealth: comments on “Salvaging the concept of nudge”. *Journal of Medical Ethics*, 39(8), 494. <https://doi.org/10.1136/medethics-2012-101109>
- Attwood, A. S., Scott-Samuel, N. E., Stothart, G., Munafò, M. R., & Campanella, S. (2012). Glass Shape Influences Consumption Rate for Alcoholic Beverages. *PLoS ONE*, 7(8), e43007. <https://doi.org/10.1371/journal.pone.0043007>
- Australian Government - Department of Health. (2016). *Post-Implementation Review Tobacco Plain Packaging 2016*. Retrieved from <http://ris.pmc.gov.au/sites/default/files/posts/2016/02/Tobacco-Plain-Packaging-PIR.pdf>
- Åvitsland, A., Solbraa, A. K., & Riiser, A. (2017). Promoting workplace stair climbing: sometimes, not interfering is the best. *Archives of Public Health*, 75, 2. <https://doi.org/10.1186/s13690-016-0170-8>
- Baer, J. S., Stacy, A., & Larimer, M. (1991). Biases in the perception of drinking norms among college students. *Journal of Studies on Alcohol*, 52(6), 580–586.
- Baldwin, R. (2014). From Regulation to Behaviour Change: Giving Nudge the Third Degree. *The Modern Law Review*, 77(6), 831–857. <https://doi.org/10.1111/1468-2230.12094>
- Baldwin, R., Cave, M., & Lodge, M. (2011). *Understanding Regulation: Theory, Strategy and Practice* (2nd ed.). Oxford: Oxford University Press.

- Barber, K., Falvey, S., Hamilton, C., Collett, D., & Rudge, C. (2006). Potential for organ donation in the United Kingdom: audit of intensive care records. *BMJ (Clinical Research Ed.)*, *332*(7550), 1124–1127. <https://doi.org/10.1136/bmj.38804.658183.55>
- Barton, A., & Grüne-Yanoff, T. (2015). From Libertarian Paternalism to Nudging—and Beyond. *Review of Philosophy and Psychology*, *6*(3), 341–359. <https://doi.org/10.1007/s13164-015-0268-x>
- Behavioural Insights Team. (2013). *Applying Behavioural Insights to Organ Donation: preliminary results from a randomised controlled trial*. Retrieved from <https://gds.blog.gov.uk/2014/03/18/organ-donor-register/>
- Benartzi, S., & Thaler, R. H. (2013). Behavioral Economics and the Retirement Savings Crisis. *Science*, *339*(6124), 1152–1153.
- Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). The Importance of default options for retirement saving outcomes: Evidence from the United States. In S. J. Kay & T. Sinha (Eds.), *Lessons from Pension Reform in the Americas* (pp. 59–87). Oxford: Oxford University Press.
- Binder, M., & Lades, L. K. (2014). *Autonomy-enhancing paternalism*.
- Blamey, A., Mutrie, N., & Aitchison, T. (1995). Health promotion by encouraged use of stairs. *BMJ*, *311*(7000), 289–290. <https://doi.org/10.1136/bmj.311.7000.289>
- Blumenthal-Barby, J. S., & Burroughs, H. (2012). Seeking Better Health Care Outcomes: The Ethics of Using the “Nudge.” *The American Journal of Bioethics*, *12*(2), 1–10. <https://doi.org/10.1080/15265161.2011.634481>
- Bonell, C., McKee, M., Fletcher, A., Wilkinson, P., & Haines, A. (2011). One nudge forward, two steps back. *BMJ*, *342*, d401–d401. <https://doi.org/10.1136/bmj.d401>
- Boreham, C. A. G., Wallace, W. F. M., & Nevill, A. (2000). Training Effects of Accumulated Daily Stair-Climbing Exercise in Previously Sedentary Young Women. *Preventive Medicine*, *30*(4), 277–281. <https://doi.org/10.1006/pmed.2000.0634>
- Borsari, B., & Carey, K. B. (2003). Descriptive and injunctive norms in college drinking: a meta-analytic integration. *Journal of Studies on Alcohol*, *64*(3), 331–341.
- Botti, S., & McGill, A. L. (2006). When Choosing Is Not Deciding: The Effect of Perceived Responsibility on Satisfaction. *Journal of Consumer Research*, *33*(2), 211–219. <https://doi.org/10.1086/506302>
- Bovens, L. (2009). The Ethics of Nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference Change : Approaches from Philosophy, Economics and Psychology* (pp. 217–

- 209). Dordrecht: Springer Netherlands.
- Bowels, S. (1998). Endogenous Preferences: The Cultural Consequences of Markets and other Economic Institutions. *Journal of Economic Literature*, XXXVI, 75–111.
- Branson, C., Duffy, B., Perry, C., & Wellings, D. (2012). *Acceptable behaviour? Public opinion on behaviour change policy*. London, UK.
- Brose, L. S., Chong, C. B., Aspinall, E., Michie, S., & McEwen, A. (2014). Effects of standardised cigarette packaging on craving, motivation to stop and perceptions of cigarettes and packs. *Psychology & Health*, 29(7), 849–860. <https://doi.org/10.1080/08870446.2014.896915>
- Brown, J. R., Farrell, A. M., & Weisbenner, S. J. (2011). *The downside of defaults*. Retrieved from www.nber.org/programs/ag/rrc/NB11-01 Brown, Farrell, Weisbenner FINAL.pdf.
- Brownell, K. D., Stunkard, A. J., & Albaum, J. M. (1980). Evaluation and modification of exercise patterns in the natural environment. *The American Journal of Psychiatry*, 137(12), 1540–1545.
- Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Jonsson, M. L., & Rahali, B. (2016). *Can Nudges Be Transparent and Yet Effective?* (No. WiSo-HH Working Paper Series No. 33).
- Bucher-Koenen, T., & Lusardi, A. (2011). Financial literacy and retirement planning in Germany. *Journal of Pension Economics and Finance*, 10(4), 565–584.
- Bullers, S., Cooper, M. L., & Russell, M. (2001). Social network drinking and adult alcohol involvement. *Addictive Behaviors*, 26(2), 181–199. [https://doi.org/10.1016/S0306-4603\(00\)00099-X](https://doi.org/10.1016/S0306-4603(00)00099-X)
- Burk, J., Huang, A. A., & Luoto, J. E. (2015). *Automatic Enrollment in Retirement Savings Vehicles - Evidence from the Health and Retirement Study* (No. WR-1117).
- Byrne, A., Blake, D., Cairns, A., & Dowd, K. (2007). Default funds in UK defined contribution pension plans. *Financial Analysts Journal*, 63(4), 4748.
- Cabinet Office Behavioural Insights Team. (2011). *Behaviour Insights Team: Annual Update 2010-11*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60537/Behaviour-Change-Insight-Team-Annual-Update_acc.pdf
- Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2009). Optimal Defaults and Active Decisions. *QJEcon*, 124(4), 1639–1674.
- Chapman, G. (2014). Judge, Nudge, Dodge: Doing What’s Good for Us. In *SJDM Presidential*

Address. Long Beach, CA.

- Choi, J. J., Laibson, D., & Madrian, B. C. (2004). Plan Design and 401(k) Savings Outcomes. *National Tax Journal*, *57*(2), 275–298.
- Cohen, I. G., Lynch, H. F., & Robertson, C. T. (2016). *Nudging Health: Health Law and Behavioral Economics*. (I. G. Cohen, H. F. Lynch, & C. T. Robertson, Eds.). Johns Hopkins University Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.
- Coleman, K. J., & Gonzalez, E. C. (2001). Promoting Stair Use in a US–Mexico Border Community. *American Journal of Public Health*, *91*(12), 2007–2009. <https://doi.org/10.2105/AJPH.91.12.2007>
- Davidai, S., Gilovich, T., & Ross, L. D. (2013). The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences*, *109*(38), 15201–15205. <https://doi.org/10.1073/pnas.1211695109>
- De Neys, W. D. (2006). Dual Processing in Reasoning: Two Systems but One Reasoner. *Psychological Science*, *17*(5), 428–433. <https://doi.org/10.1111/j.1467-9280.2006.01723.x>
- DeJong, W., Franz, H. G., Wolfe, S. M., Nathan, H., Payne, D., Reitsma, W., & Beasley, C. (1998). Requesting organ donation: an interview study of donor and nondonor families. *American Journal of Critical Care*, *7*(1), 13–23.
- Den Hartogh, G. (2012). The role of the relatives in opt-in systems of postmortal organ procurement. *Medline, Health Care, and Philosophy*, *15*(2), 195–205. [https://doi.org/DenHartogh, G. \(2012\). The role of the relatives in opt-in systems of postmortal organ procurement. Medicine, Health Care, and Philosophy, 15\(2\), 195–205. <http://doi.org/10.1007/s11019-011-9317-0>](https://doi.org/DenHartogh, G. (2012). The role of the relatives in opt-in systems of postmortal organ procurement. Medicine, Health Care, and Philosophy, 15(2), 195–205. http://doi.org/10.1007/s11019-011-9317-0)
- Department for Work & Pensions. (2008). Pension Act 2008. Retrieved from <http://www.legislation.gov.uk/ukpga/2008/30/contents>
- Department of Health & Social Care. (2017). *Consultation on introducing ‘opt-out’ consent for organ and tissue donation in England*. Retrieved from <https://www.gov.uk/government/consultations/introducing-opt-out-consent-for-organ-and-tissue-donation-in-england/consultation-on-introducing-opt-out-consent-for-organ-and-tissue-donation-in-england>
- Downs, J. S., Loewenstein, G., & Wisdom, J. (2009). Strategies for Promoting Healthier Food

- Choices. *American Economic Review*, 99(2), 159–164.
<https://doi.org/10.1257/aer.99.2.159>
- Dunlop, S. M., Dobbins, T., Young, J. M., Perez, D., & Currow, D. C. (2014). Impact of Australia's introduction of tobacco plain packs on adult smokers' pack-related perceptions and responses: results from a continuous tracking survey. *BMJ Open*, 4(12), e005836. <https://doi.org/10.1136/bmjopen-2014-005836>
- Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge, UK: Cambridge University Press.
- Dworkin, G. (2012). Lying and nudging. *Journal of Medical Ethics*, 39(8), 496–497.
- Elbel, B., Gyamfi, J., & Kersh, R. (2011). Child and adolescent fast-food choice and the influence of calorie labeling: a natural experiment. *International Journal of Obesity (London)*, 35(4), 493–500. <https://doi.org/10.1038/ijo.2011.4>
- Emrich, T. E., Qi, Y., Lou, W. Y., L'Abbe, M. R., Garcia, R., & Cavalcanti, R. (2017). Traffic-light labels could reduce population intakes of calories, total fat, saturated fat, and sodium. *PLOS ONE*, 12(2), e0171188. <https://doi.org/10.1371/journal.pone.0171188>
- Ennett, S. T., Bauman, K. E., Hussong, A., Faris, R., Foshee, V. A., Cai, L., & DuRant, R. H. (2006). The Peer Context of Adolescent Substance Use: Findings from Social Network Analysis. *Journal of Research on Adolescence*, 16(2), 159–186. <https://doi.org/10.1111/j.1532-7795.2006.00127.x>
- Epstein, S. (1973). The self-concept revisited: Or a theory of a theory. *American Psychologist*, 28(5), 404–416. <https://doi.org/10.1037/h0034679>
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *The American Psychologist*, 49, 729–724.
- Eurobarometer. (2007). *Europeans and organ donation report*.
- Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. *International Journal of Psychology* (Vol. 31). East Sussex: Psychology Press.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Exley, M., White, N., & Martin, J. H. (2002). Why families say no to organ donation. *Critical Care Nurse*, 22(6), 44–51.
- Fardy, P. S., & Ilmarinen, J. (1975). Evaluating the effects and feasibility of an at work stairclimbing intervention program for men. *Medicine and Science in Sports*, 7(2), 91–93.

- Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: public attitudes towards overt and covert nudges. *Judgement and Decision Making*, 8(3), 202. <https://doi.org/10.1177/2332858416674007>
- Felsen, G., & Reiner, P. B. (2015). What can neuroscience contribute to the debate over nudging? *Review of Philosophy and Psychology*, 6(3), 469–479. <https://doi.org/10.1007/s13164-015-0240-9>
- Financial Services and the Treasury Bureau. (2015). *Mandatory Provident Fund Schemes (Amendment) Bill 2015*. HK.
- Finkelstein, E. A., Strombotne, K. L., Chan, N. L., & Krieger, J. (2011). Mandatory Menu Labeling in One Fast-Food Chain in King County, Washington. *American Journal of Preventive Medicine*, 40(2), 122–127. <https://doi.org/10.1016/j.amepre.2010.10.019>
- Foxcraft, D. R., Moreira, M. T., Santimano, N. M. L., & Smith, L. A. (2015). Social norms information for alcohol misuse in university and college students (Review). *Cochrane Database of Systematic Reviews*, (12). <https://doi.org/10.1002/14651858.CD006748.pub4>
- Ghorbani, F., Khoddami-Vishteh, H. R., Ghobadi, O., Shafaghi, S., Louyeh, A. R., & Najafizadeh, K. (2011). Causes of family refusal for organ donation. *Transplantation Proceedings*, 43(2), 405–406. <https://doi.org/10.1016/j.transproceed.2011.01.031>
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4), 684–704. <https://doi.org/http://dx.doi.org/10.1037/0033-295X.102.4.684>
- Gigerenzer, G., & Regier, T. (1996). How Do We Tell an Association From a Rule? Comment on Sloman (1996). *Psychological Bulletin*, 119(1), 23–26.
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.
- Goldlin, J. (2015). Which Way To Nudge? Uncovering Preferences in the Behavioral Age. *The Yale Law Journal*, 125(1), 226–270.
- Gomberg, L., Schneider, S. K., & DeJong, W. (2001). Evaluation of a social norms marketing campaign to reduce high-risk drinking at the University of Mississippi. *The American Journal of Drug and Alcohol Abuse: Encompassing All Addictive Disorders*, 27(2), 375–389. <https://doi.org/10.1081/ADA-100103715>
- Gomberg, L., Schneider, S. K., & DeJong, W. (2001). Evaluation of a social norms marketing

- campaign to reduce high-risk drinking at the University of Mississippi*. *The American Journal of Drug and Alcohol Abuse: Encompassing All Addictive Disorders*, 27(2), 375–389.
- Goodwin, T. (2012). Why we should reject “nudge.” *Politics*, 32(2), 85–92. <https://doi.org/10.1111/j.1467-9256.2012.01430.x>
- Gorton, D., Ni Mhurchu, C., Chen, M., & Dixon, R. (2009). Nutrition labels: a survey of use, understanding and preferences among ethnically diverse shoppers in New Zealand. *Public Health Nutrition*, 12(09), 1359. <https://doi.org/10.1017/S1368980008004059>
- Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, 32(3), 463–483. <https://doi.org/10.1017/S0266267115000425>
- Grüne-Yanoff, T., & Hertwig, R. (2015). Nudge Versus Boost: How Coherent are Policy and Theory? *Minds and Machines*, 26(1–2), 1–52. <https://doi.org/10.1007/s11023-015-9367-9>
- Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public Views on Policies Involving Nudges. *Review of Philosophy and Psychology*, 6(3), 439–453. <https://doi.org/10.1007/s13164-015-0263-2>
- Hahn, U. (2011). The problem of circularity in evidence, argument, and explanation. *Perspectives on Psychological Science*, 6(2), 172–182.
- Haines, M. P., Barker, G., & Rice, R. (2003). The NIU Experiment: A case study of the social norms approach. In H. W. Perkins (Ed.), *The Social Norms Approach to Preventing School and College Age Substance Abuse: A Handbook for Educators, Counselors, and Clinicians* (p. 22). San Francisco, Calif: Jossey-Bass.
- Haines, M. P., & Spear, S. F. (1996). Changing the perception of the norm: a strategy to decrease binge drinking among college students. *Journal of American College Health : J of ACH*, 45(3), 134–140. <https://doi.org/10.1080/07448481.1996.9936873>
- Hall, M. G., Marteau, T. M., Sunstein, C. R., Ribisl, K. M., Noar, S. M., Orlan, E. N., & Brewer, N. T. (2018). Public support for pictorial warnings on cigarette packs: an experimental study of US smokers. *J Behav Med*. <https://doi.org/10.1007/s10865-018-9910-2>
- Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., Ekelund, U., & Group., L. P. A. S. W. G. (2012). Global physical activity levels: surveillance progress, pitfalls, and prospects. *The Lancet*, 380(9838), 247–257.
- Hammond, D., Daniel, S., & White, C. M. (2013). The effect of cigarette branding and plain packaging on female youth in the United Kingdom. *The Journal of Adolescent Health* :

- Official Publication of the Society for Adolescent Medicine*, 52(2), 151–157.
<https://doi.org/10.1016/j.jadohealth.2012.06.003>
- Hammond, D., Fong, G. T., McNeill, A., Borland, R., Cummings, K. M., & Hammond, D. (2006). Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: findings from the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control*, 15, 19–25. <https://doi.org/10.1136/tc.2005.012294>
- Hammond, D., & Parkinson, C. (2009). The impact of cigarette package design on perceptions of risk. *Journal of Public Health*, 31(3), 345–353. <https://doi.org/10.1093/pubmed/fdp066>
- Hancock, L., Abhold, J., Gascoigne, J., & Altekruse, M. (2002). Applying Social Norms Marketing to Tobacco Cessation and Prevention: Lessons Learned from Three Campaigns. *The Report on Social Norms, Working Pa.* Retrieved from https://www.mendeley.com/research/applying-social-norms-marketing-tobacco-cessation-prevention-lessons-learned-three-campaigns/?utm_source=desktop&utm_medium=1.15&utm_campaign=open_catalog&userDocumentId=%7B49f39eea-2ba5-40fa-b274-05ed28604d4d%7D
- Hancock, L., & Henry, N. (2003). *Perceptions, Norms and Tobacco Use in College Residence Hall Freshmen: Evaluation of a Social Norms Marketing Intervention*. San Francisco: Jossery-Bass.
- Hanna, S. D., Kim, K. T., & Chen, S. C. C. (2016). Retirement savings. In J. J. Xiao (Ed.), *Handbook of Consumer Finance Research* (pp. 33–43). New York: Springer International Publishing.
- 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*, 1, 3–28.
- Harnack, L. J., & French, S. A. (2008). Effect of point-of-purchase calorie labeling on restaurant and cafeteria food choices: A review of the literature. *International Journal of Behavioral Nutrition and Physical Activity*, 5(1), 51. <https://doi.org/10.1186/1479-5868-5-51>
- Harris, A. J. L., Corner, A., & Hahn, U. (2013). James is polite and punctual (and useless): A Bayesian formalisation of faint praise. *Thinking & Reasoning*, 19(3–4), 414–429. <https://doi.org/10.1080/13546783.2013.801367>
- Hausman, D. M., & Welch, B. (2010). Debate: To Nudge or Not to Nudge. *The Journal of Political Philosophy*, 18(1), 123–136.

- Hedlin, S., & Sunstein, C. R. (2016). Does Active Choosing Promote Green Energy Use: Experimental Evidence. *Ecology*, *43*, 107.
- Hogg, M. (2011). *Do we need a deeper, more complex conversation with the public about global issues? A review of literature.*
- Holden, S. S., Zlatevska, N., & Dubelaar, C. (2016). Whether Smaller Plates Reduce Consumption Depends on Who's Serving and Who's Looking: A Meta-Analysis. *Journal of the Association for Consumer Research*, *1*(1), 134–146.
- Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Kelly, M. P., Nakamura, R., ... Ogilvie, D. (2013). Altering micro-environments to change population health behaviour: towards an evidence base for choice architecture interventions. *BMC Public Health*, *13*(1), 1218. <https://doi.org/10.1186/1471-2458-13-1218>
- Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Lewis, H. B., Wei, Y., ... Ogilvie, D. (2015). Portion, package or tableware size for changing selection and consumption of food, alcohol and tobacco. In G. J. Hollands (Ed.), *Cochrane Database of Systematic Reviews*. UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/14651858.CD011045.pub2>
- Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Kelly, M. P., Nakamura, R., ... Ogilvie, D. (2013). *Altering choice architecture to change population health behaviour: a large-scale conceptual and empirical scoping review of interventions within micro-environments*. Cambridge.
- Holmes, A. S., Serrano, E. L., Machin, J. E., Duetsch, T., & Davis, G. C. (2013). Effect of different children's menu labeling designs on family purchases. *Appetite*, *62*, 198–202. <https://doi.org/10.1016/j.appet.2012.05.029>
- Horack, S., & Wood, A. (2005). *An evaluation of scheme joining techniques in workplace pension schemes with an employer contribution: Case Study. Research Report (Vol. 292)*. London.
- House of Lords. (2011). *Behaviour Change*. London. Retrieved from <http://www.parliament.uk/business/lords>
- Hull, C. L. (1943). *Principles of behaviour: An introduction to behaviour theory*. New York: Appleton-Century Crofts.
- Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2016). When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects. *Available at SSRN*.
- Jog, V., & Lee, I. (2016). Reforming Canada's retirement savings system-solutions for a non-

- existent problem? *Public Finance & Management*, 16(2).
- John, P., Moseley, A., Wales, C., Smith, G., Richardson, L., Stoker, G., & Cotterill, S. (2013). *Nudge, Nudge, Think, Think: Experimenting with Ways to Change Civic Behaviour*. London: Bloomsbury Academic.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults, Framing and Privacy: Why Opting In-Opting Out. *Marketing Letters*, 13(1), 5–15.
- Johnson, E. J., & Goldstein, D. G. (2003). Do defaults save lives? *Science*, 302(5649), 1338–1339. <https://doi.org/10.1126/science.1091721>
- Jolls, C., Sunstein, C. R., & Thaler, R. (1998). *A Behavioral Approach to Law and Economics*. *Stanford Law Review*. Retrieved from <https://ssrn.com/abstract=74927>
- Jung, J. Y., & Mellers, B. A. M. (2016). American attitudes toward nudges. *Judgment and Decision Making*, 11(1), 62–74.
- Junghans, A. F., Cheung, T. T. L., & De Ridder, D. T. D. (2015). Under consumers' scrutiny - an investigation into consumers' attitudes and concerns about nudging in the realm of health behavior. *BMC Public Health*, 15, 336. <https://doi.org/10.1186/s12889-015-1691-8>
- Kahneman, D. (2003). Maps of bounded rationality: a perspective on intuitive judgement and choice. In T. Frangmyr (Ed.), *Lex Prix Nobel: The Nobel Prizes 2002* (pp. 449–489). Stockholm: Almqvist & Wiksell International.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49–81). New York: Cambridge University Press.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoack & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267–293). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 493–508). New York: Cambridge University Press. [https://doi.org/10.1016/0010-0277\(82\)90022-1](https://doi.org/10.1016/0010-0277(82)90022-1)
- Kerr, J., Eves, F., & Carroll, D. (2001a). Six-Month Observational Study of Prompted Stair Climbing. *Preventive Medicine*, 33(5), 422–427. <https://doi.org/10.1006/pmed.2001.0908>
- Kerr, J., Eves, F. F., & Carroll, D. (2001b). Encouraging stair use: Stair riser banners are better

- than poster prompts. *American Journal of Public Health*, 91(8), 1192–1193. Retrieved from https://www.mendeley.com/research/encouraging-stair-stair-riser-banners-better-poster-prompts/?utm_source=desktop&utm_medium=1.15&utm_campaign=open_catalog&userDocumentId=%7Be50beb00-7aee-4cfa-a07e-eb2b0e721b86%7D
- Kerr, J., Eves, F. F., & Carroll, D. (2001c). Getting more people on the stairs: the impact of a new message format. *Journal of Health Psychology*, 6(5), 495–500. <https://doi.org/10.1177/135910530100600503>
- Kiszko, K. M., Martinez, O. D., Abrams, C., & Elbel, B. (2014). The Influence of Calorie Labeling on Food Orders and Consumption: A Review of the Literature. *Journal of Community Health*, 39(6), 1248–1269. <https://doi.org/10.1007/s10900-014-9876-0>
- Klaczynski, P. A. (2001). Analytic and Heuristic Processing Influences on Adolescent Reasoning and Decision Making. *Child Development*, 72(3), 844–861.
- Knoef, M., Been, J., Alessie, R., Caminada, K., Goudswaard, K., & Kalwij, A. (2016). Measuring retirement savings adequacy: developing a multi-pillar approach in the Netherlands. *Journal of Pension Economics and Finance*, 15(1), 55–89.
- Kosters, M., & Van der Heijden, J. (2015). From mechanism to virtue: Evaluating Nudge theory. *Evaluation*, 21(3), 276–291. <https://doi.org/10.1177/1356389015590218>
- Kruglanski, A. W. (2013). Only One? The Default Interventionist Perspective as a Unimodel-Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, 8(3), 242–247. <https://doi.org/10.1177/1745691613483477>
- Lagnado, D. A., Fenton, N., & Neil, M. (2012). Legal idioms: a framework for evidential reasoning. *Argument & Computation*, 4(1), 46–63. <https://doi.org/10.1080/19462166.2012.682656>
- Laplace, S. (1951). *A philosophical essay on probabilities (In F.W. Truscott & F. L. Emory Trans Ed.)*. New York: Dover (Original Work Published 1814).
- Laszlo, E. (1989). *The Inner Limits of Mankind*. London: One World.
- Lee, K. K., Perry, A. S., Wolf, S. A., Agarwal, R., Rosenblum, R., Fischer, S., ... Silver, L. D. (2012). Promoting Routine Stair Use Evaluating the Impact of a Stair Prompt Across Buildings. *American Journal of Preventive Medicine*, 42(2), 136–141. <https://doi.org/10.1016/j.amepre.2011.10.005>
- Lewis, A., & Eves, F. (2012). Prompt before the choice is made: Effects of a stair-climbing intervention in university buildings. *British Journal of Health Psychology*, 17(3), 631–

643. <https://doi.org/10.1111/j.2044-8287.2011.02060.x>
- Lewis, M. A., & Neighbors, C. (2006). Social norms approaches using descriptive drinking norms education: a review of the research on personalized normative feedback. *Journal of American College Health, 54*(4), 213–218. <https://doi.org/10.3200/JACH.54.4.213-218>
- Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, Effectiveness, and Ethics. *Basic and Applied Social Psychology, 39*(6), 293–306. <https://doi.org/https://doi.org/10.1080/01973533.2017.1356304>
- Lin, Y., Osman, M., Harris, A. J. L., & Read, D. (2018). Underlying Wishes and Nudged Choices. *JEP: Applied*.
- Linkenbach, J., Perkins, H. W., & DeJong, W. (2003). *Parent's Perceptions of Parenting Norms: Using the Social Norms Approach to Reinforce Effective Parenting*. San Francisco: Jossery-Bass.
- Liu, P. J., Wisdom, J., Roberto, C. A., Liu, L. J., & Ubel, P. A. (2014). Using Behavioral Economics to Design More Effective Food Policies to Address Obesity. *Applied Economic Perspectives and Policy, 36*(1), 6–24. <https://doi.org/10.1093/aep/ppt027>
- Loewenstein, G., Asch, D. A., Friedman, J. Y., Melichar, L. A., Volpp, K. G., Loewenstein, G., ... Volpp, K. G. (2012). Can behavioural economics make us healthier? *BMJ, 344*, e3482. <https://doi.org/10.1136/bmj.e3482>
- Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2014). Warning: You are About to Be Nudged. *SSRN Electronic Journal, 1*–23. <https://doi.org/10.2139/ssrn.2417383>
- Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2015). Warning: You are about to be nudged. *Behavioral Science & Policy, 1*(1), 35–42.
- Loosemore, T. (2014). One link on GOV.UK - 350,000 more organ donors. Retrieved from <https://gds.blog.gov.uk/2014/03/18/organ-donor-register/>
- Lusardi, A., & Mitchell, O. S. (2011a). Financial literacy and planning: Implications for retirement wellbeing (No. w17078). In A. Lusardi & O. S. Mitchell (Eds.), *Financial Literacy. Implications for Retirement Security and the Financial Marketplace* (pp. 17–39). New York: Oxford University Press.
- Lusardi, A., & Mitchell, O. S. (2011b). Financial literacy around the world: an overview. *Journal of Pension Economics and Finance, 10*(4), 497–508.
- Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature, 52*(1), 5–44.
- Magnusson, R. S. (2009). Rethinking global health challenges: towards a 'global compact' for

- reducing the burden of chronic disease. *Public Health*, 123(3), 265–274.
- Mandell, L., & Klein, L. S. (2007). Motivation and financial literacy. *Financial Services Review*, 16(2), 105.
- Marshall, A. L., Bauman, A. E., Patch, C., Wilson, J., & Chen, J. (2002). Can motivational signs prompt increases in incidental physical activity in an Australian health-care facility? *Health Education Research*, 17(6), 743–749. <https://doi.org/10.1093/her/17.6.743>
- Marteau, T. M., Ogilvie, D., Roland, M., Suhrcrke, M., & Kelly, M. P. (2011). Judging nudging: can nudging improve population health? *BMJ (Clinical Research Ed.)*, 342, d228. <https://doi.org/10.1136/bmj.d228>
- Maynard, O. M., Munafò, M. R., & Leonards, U. (2013). Visual attention to health warnings on plain tobacco packaging in adolescent smokers and non-smokers. *Addiction*, 108(2), 413–419. <https://doi.org/10.1111/j.1360-0443.2012.04028.x>
- Mazzocchi, M., Cagnone, S., Bech-Larsen, T., Niedźwiedzka, B., Saba, A., Shankar, B., ... Traill, W. B. (2015). What is the public appetite for healthy eating policies? Evidence from a cross-European survey. *Health Economics, Policy and Law*, 10(03), 267–292. <https://doi.org/10.1017/S1744133114000346>
- McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, 17(5), 414–420. <https://doi.org/10.1111/j.1467-9280.2006.01721.x>
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(42), 1–11.
- Mill, J. S. (1863). *On Liberty*. Boston, MA: Ticknor and Fields.
- Mills, C. (2013). Why nudges matter: a reply to Goodwin. *Politics*, 33(1), 28–36. <https://doi.org/10.1111/j.1467-9256.2012.01450.x>
- 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), 195. <https://doi.org/10.1111/1475-6765.12073>
- Moodie, C., & Mackintosh, A. M. (2013). Young adult women smokers' response to using plain cigarette packaging: a naturalistic approach. *BMJ Open*, 3(3), e002402. <https://doi.org/10.1136/bmjopen-2012-002402>
- Moodie, C., Mackintosh, A. M., Hastings, G., & Ford, A. (2011). Young adult smokers' perceptions of plain packaging: a pilot naturalistic study. *Tobacco Control*, 20(5), 367–

373. <https://doi.org/10.1136/tc.2011.042911>
- Moodie, C., Stead, M., Bauld, L., McNeill, A., Angus, K., Hinds, K., ... O'Mara-Eves, A. (2012). *Plain Tobacco Packaging: A Systematic Review*. London. Retrieved from http://dspace.stir.ac.uk/bitstream/1893/24420/1/Moodie_et_al_2012_PHRC_PlainPacks.pdf
- Mossialos, E., Costa-Font, J., & Rudisill, C. (2008). Does organ donation legislation affect individuals' willingness to donate their own or their relative's organs? Evidence from European Union survey data. *BMC Health Services Research*, 8(1), 48. <https://doi.org/10.1186/1472-6963-8-48>
- Munafò, M. R., Roberts, N., Bauld, L., & Leonards, U. (2011). Plain packaging increases visual attention to health warnings on cigarette packs in non-smokers and weekly smokers but not daily smokers. *Addiction*, 106(8), 1505–1510. <https://doi.org/10.1111/j.1360-0443.2011.03430.x>
- Mutran, E. J., Reitzes, D. C., & Fernandez, M. E. (1997). Factors that influence attitudes toward retirement. *Research on Aging*, 19(3), 251–273.
- National Advisory Committee on Bioethics. (2015). *Nudging in Public Health – An Ethical Framework*.
- NHSBT. (2015). News Release: Call for everyone in the UK to discuss organ donation as the number of deceased organ donors in the UK drops for the first time in more than a decade. Retrieved December 4, 2015, from http://www.nhsbt.nhs.uk/news-and-media/news-articles/news_2015_07_20.asp
- NHSBT. (2016). Highest number of organ transplants ever across UK, but many families still say no to donation. Retrieved from <https://www.organdonation.nhs.uk/news-and-campaigns/news/highest-number-of-organ-transplants-ever-across-uk-but-many-families-still-say-no-to-donation/>
- NHSBT. (2017). Organ Donation and Transplantation: Activity Report 2016/17. Retrieved from https://nhsbt.dbe.blob.core.windows.net/umbraco-assets-corp/4657/activity_report_2016_17.pdf
- Nikolaou, C. K., Hankey, C. R., & Lean, M. E. J. (2014). Preventing weight gain with calorie-labelling. *Obesity*, 22(11), 2277–2283.
- Nishida, Y., Akaoka, I., & Nishizawa, T. (1975). Effect of sex hormones on uric acid metabolism in rats. *Experientia*, 31(10), 1134–1135. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1204720>

- Nocon, M., Müller-Riemenschneider, F., Nitzschke, K., & Willich, S. N. (2010). Review Article: Increasing physical activity with point-of-choice prompts--a systematic review. *Scandinavian Journal of Public Health*, 38(6), 633–638. <https://doi.org/10.1177/1403494810375865>
- Nomura, T., Yoshimoto, Y., Akezaki, Y., & Sato, A. (2009). Changing behavioral patterns to promote physical activity with motivational signs. *Environmental Health and Preventive Medicine*, 14(1), 20–25. <https://doi.org/10.1007/s12199-008-0053-x>
- Noone, J. H. (2010). *Psychological and socioeconomic factors influencing men and women's planning for retirement: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Psychology at Massey University, Turitea, New Zealand (Doct.*
- Nørnberg, T. R., Houlby, L., Skov, L., & Pérez-Cueto, F. J. (2016). Choice architecture interventions for increased vegetable intake and behaviour change in a school setting: a systematic review. *Perspect Public Health*, 136(3), 134–142. <https://doi.org/10.1177/1757913915596017>
- OECD. (2013). *Pensions at a glance 2013: OECD and G20 indicators*. [https://doi.org/OECD\(2013\), Pensions at a Glance 2013: OECD and G20 Indicators](https://doi.org/OECD(2013),Pensions%20at%20a%20Glance%202013%3A%20OECD%20and%20G20%20Indicators), OECD Publishing. http://dx.doi.org/10.1787/pension_glance-2013-en
- Office of Fair Trading. (2013). *Defined contribution workplace pension market study*. Retrieved from http://webarchive.nationalarchives.gov.uk/20131101164215/http://www.oft.gov.uk/shared_of/market-studies/oft1505;jsessionid=BA96A643DCEA2A81D543CF38CB586719
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, 12(6), 237–241. <https://doi.org/10.1016/j.tics.2008.02.014>
- Organdonor.gov. (2017). The Need is Real: Data. Retrieved February 20, 2016, from <http://www.organdonor.gov/about/data.html>
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11(6), 988–1010. <https://doi.org/10.3758/BF03196730>
- Osman, M. (2007). Can tutoring improve performance on a reasoning task under deadline conditions? *Memory & Cognition*, 35(2), 342–351.
- Osman, M. (2013). A Case Study: Dual-Process Theories of Higher Cognition-Commentary on Evans & Stanovich (2013). *Prospect Psychological Science*, 8(3), 248–252. <https://doi.org/10.1177/1745691613483475>

- Osman, M. (2014a). *Future-minded: The Psychology of Agency and Control*. New York: Palgrave Macmillan.
- Osman, M. (2014b). What are the essential cognitive requirements for prospection (thinking about the future)? *Frontiers in Psychology*, 5, 626. <https://doi.org/10.3389/fpsyg.2014.00626>
- Osman, M. (2016). Nudge: How Far Have We Come? *Æconomia. History, Methodology, Philosophy*, (6–4), 557–570. Retrieved from <https://oeconomia.revues.org/2490>
- Osman, M., & Stavy, R. (2006). Intuitive rules: From formative to developed reasoning. *Psychonomic Bulletin & Review*, 13(935–53).
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.
- Pearson, M., & West, P. (2003). Drifting smoke rings: Social network analysis and Markov processes in a longitudinal study of friendship groups and risk-taking. *Connections*, 25(2), 59–76.
- Pechey, R., Burge, P., Mentzakis, E., Suchrcke, M., & Marteau, T. M. (2014). Public acceptability of population-level interventions to reduce alcohol consumption: A discrete choice experiment. *Social Science & Medicine*, 113, 104–109.
- Perkins, H. W., Meilman, P. W., Leichter, J. S., Cashin, J. R., & Presley, C. A. (1999). Misperceptions of the Norms for the Frequency of Alcohol and Other Drug Use on College Campuses. *Journal of American College Health*, 47(6), 253–258. <https://doi.org/10.1080/07448489909595656>
- 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.
- Puig-Ribera, A., & Eves, F. F. (2010). Promoting stair climbing in Barcelona: similarities and differences with interventions in English-speaking populations. *European Journal of Public Health*, 20(1), 100–102. <https://doi.org/10.1093/eurpub/ckp059>
- Radecki, C. M., & Jaccard, J. (1997). Psychological aspects of organ donation: a critical review and synthesis of individual and next-of-kin donation decisions. *Health Psychology*, 16(2), 183–195.
- Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, parallel constraint satisfaction processes, and gestalt principles: (Re)introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review*, 1, 26–53.

- Rebonato, R. (2013). A Critical Assessment of Libertarian Paternalism.
- Reisch, L. A., & Sunstein, C. R. (2016). Do Europeans like nudges? *Judgement and Decision Making*, *11*(4), 310–325.
- Reisch, L. A., Sunstein, C. R., & Gwozdz, W. (2016). Better Than A Whip? European Attitudes Toward Health Nudges. *Food Policy*, *69*, 1–10. <https://doi.org/doi.org/10.1016/j.foodpol.2017.01.007>
- Rithalia, A., McDaid, C., Suekarran, S., Norman, G., Myers, L., & Sowden, A. (2009). A systematic review of presumed consent systems for deceased organ donation. *Health Technology Assessment*, *13*(26). <https://doi.org/10.3310/hta13260>
- Robinson, E., Nolan, S., Tudur-Smith, C., Boyland, E. J., Harrold, J. A., Hardman, C. A., & Halford, J. C. G. (2014). Will smaller plates lead to smaller waists? A systematic review and meta-analysis of the effect that experimental manipulation of dishware size has on energy consumption. *Obesity Reviews*, *15*(10), 812–821. <https://doi.org/10.1111/obr.12200>
- Rosenblum, A. M., Horvat, L. D., Siminoff, L. A., Prakash, V., Beitel, J., & Garg, A. X. (2012). The authority of next-of-kin in explicit and presumed consent systems for deceased organ donation: an analysis of 54 nations. *Nephrology Dialysis Transplantation*, *27*(6), 2533–2546. <https://doi.org/10.1093/ndt/gfr619>
- Russell, W. D., Dzewaltowski, D. A., & Ryan, G. J. (1999). The Effectiveness of a Point-of-Decision Prompt in Deterring Sedentary Behavior. *American Journal of Health Promotion*, *13*(5), 257–259. <https://doi.org/10.4278/0890-1171-13.5.257>
- Sacks, G., Rayner, M., & Swinburn, B. (2009). Impact of front-of-pack “traffic-light” nutrition labelling on consumer food purchases in the UK. *Health Promotion International*, *24*(4), 344–352. <https://doi.org/10.1093/heapro/dap032>
- Saghai, Y. (2013). Salvaging the concept of nudge. *Journal of Medical Ethics*, *39*(8), 487–493. <https://doi.org/10.1136/medethics-2012-100727>
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*(3), 414.
- Schwarz, N. (1998). Warmer and More Social: Recent Developments in Cognitive Social Psychology. *Annual Review of Sociology*, *24*(1), 239–264. <https://doi.org/10.1146/annurev.soc.24.1.239>
- Science and Technology Committee. (2014). *Letter to the Houe of Lords*. London. Retrieved

- from <http://www.parliament.uk/documents/lords-committees/science-technology/behaviourchangefollowup/LetwinBehaviourChangeLtr20140722.pdf>
- Selinger, E., & White, K. P. (2012). What Counts as a Nudge? *The American Journal of Bioethics*, *12*(2), 11–12. <https://doi.org/10.1080/15265161.2011.634485>
- Selinger, E., & Whyte, K. P. (2012). Nudging cannot solve complex policy problems. *European Journal of Risk Regulation*, *3*(1), 26–31.
- Shepherd, L., O’Carroll, R. E., & Ferguson, E. (2014). An international comparison of deceased and living organ donation/transplant rates in opt-in and opt-out systems: a panel study. *BMC Medicine*, *12*(1), 131. <https://doi.org/10.1186/s12916-014-0131-4>
- Shu, S. B., Bang, M. H., & Weber, E. U. (2016). Informed Consent to Choice Architecture: Transparency, Trust, and Perceived Effectiveness. *In Preparation*.
- Shu, S. B., & Payne, J. (2015). Psychological Factors in Savings and Decumulation Decisions. *NA-Advances in Consumer Research*, *43*.
- Siminoff, L. A., & Lawrence, R. H. (2002). Knowing Patients’ Preferences about Organ Donation: Does it Make a Difference? *The Journal of Trauma*, *53*(4), 754–760. <https://doi.org/10.1097/01.TA.0000030052.69240.4F>
- Simon, D., & Holyoak, K. J. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review*, *6*, 283–294.
- Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814–837.
- Sinclair, S. E., Cooper, M., & Mansfield, E. D. (2014). The Influence of Menu Labeling on Calories Selected or Consumed: A Systematic Review and Meta-Analysis. *Journal of the Academy of Nutrition and Dietetics*, *114*(9), 1375–1388. <https://doi.org/10.1016/j.jand.2014.05.014>
- Skov, L. R., Lourenço, S., Hansen, G. L., Mikkelsen, B. E., & Schofield, C. (2013). Choice architecture as a means to change eating behaviour in self-service settings: a systematic review. *Obesity Reviews*, *14*(3), 187–196. <https://doi.org/10.1111/j.1467-789X.2012.01054.x>
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22. <https://doi.org/10.1037/0033-2909.119.1.3>
- Soler, R. E., Leeks, K. D., Buchanan, L. R., Brownson, R. C., Heath, G. W., Hopkins, D. H., & the Task Force on Community Preventive Services. (2010). Point-of-Decision Prompts

- to Increase Stair Use A Systematic Review Update. *American Journal of Preventive Medicine*, 38(2 Suppl), S292–S300. <https://doi.org/10.1016/j.amepre.2009.10.028>
- Song, C. (2015). Financial illiteracy and pension contributions: A field experiment on compound interest in China. Retrieved from <http://dx.doi.org/10.2139/ssrn.2580856>.
- Sonnenberg, L., Gelsomin, E., Levy, D. E., Riis, J., Barraclough, S., & Thorndike, A. N. (2013). A traffic light food labeling intervention increases consumer awareness of health and healthy choices at the point-of-purchase. *Preventive Medicine*, 57(4), 253–257. <https://doi.org/10.1016/j.ypmed.2013.07.001>
- Spital, A. (1993). A consent for organ donation: Time for change. *Clinical Transplant*, 7, 525–528.
- Spital, A. (1995). Mandated choice: A plan to increase public commitment to organ donation. *JAMA*, 273(6), 504–506. <https://doi.org/10.1001/jama.1995.03520300078045>
- Spital, A. (1996). Mandated choice for organ donation: Time to give it a try. *Annals of Internal Medicine*, 125(1), 66–69.
- Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, 23(5), 701–717. <https://doi.org/10.1017/S0140525X00623439>
- Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychological Review*, 112(1), 159.
- Sunstein, C. R. (1993). Endogenous Preferences, Environmental Law. *The Journal of Legal Studies*, 22(2), 217–254.
- Sunstein, C. R. (2013). *Simple: The Future of Government*. New York: Simon & Schuster.
- Sunstein, C. R. (2014a). Nudging: A Very Short Guide. *Journal of Consumer Policy*, 37, 1–8.
- Sunstein, C. R. (2014b). *Why Nudge? The Politics of Libertarian Paternalism*. New Haven: Yale University Press.
- Sunstein, C. R. (2015a). *Choosing not to choose: Understanding the value of choice*. New York, NY: Oxford University Press.
- Sunstein, C. R. (2015b). Nudges, Agency, and Abstraction: A Reply to Critics. *Review of Philosophy and Psychology*, 6(3), 511–529. <https://doi.org/10.1007/s13164-015-0266-z>
- Sunstein, C. R. (2015c). The Ethics of Nudging. *Yale Journal on Regulation*, 32(2), 413–450.
- Sunstein, C. R. (2016a). Autonomy by Default. *The American Journal of Bioethics*, 16(11), 1–2. <https://doi.org/https://doi.org/10.1080/15265161.2016.1222169>
- Sunstein, C. R. (2016b). Do People Like Nudges? *Administrative Law Review*, Forthcoming. <https://doi.org/10.2139/ssrn.2604084>

- Sunstein, C. R. (2016c). People prefer system 2 nudges (kind of). *Duke Law Journal*, 66, 121–168.
- Sunstein, C. R. (2016d). *The Ethics of Influence: Government in the Age of Behavioral Science* (ed.). Cambridge University Press.
- Sunstein, C. R., Reisch, L. A., & Rauber, J. (2017). *A World-Wide Consensus on Nudging? Not Quite, But Almost. Regulation and Governance*. <https://doi.org/10.2139/ssrn.2955693>
- Sunstein, C. R., Reisch, L. A., & Reauber, J. (2017). A worldwide consensus on nudging? Not quite, but almost. *Regulation & Governance*, 12, 3–22.
- Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 70(4), 1159–1202.
- Swartz, J. J., Braxton, D., & Viera, A. J. (2011). Calorie menu labeling on quick-service restaurant menus: an updated systematic review of the literature. *International Journal of Behavioural Nutrition and Physical Activity*, 8, 135. <https://doi.org/10.1186/1479-5868-8-135>
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian Paternalism. *American Economic Review*, 93(2), 175–179.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.
- The Guardian. (2015). Organ donation rates held back by families refusing consent, study finds. Retrieved from <https://www.theguardian.com/australia-news/2015/mar/01/organ-donation-rates-held-back-by-families-refusing-consent-study-finds>
- Thombs, D. L., Dotterer, S., Olds, R. S., Sharp, K. E., & Raub, C. G. (2004). A close look at why one social norms campaign did not reduce student drinking. *Journal of American College Health*, 53(2), 61–68. <https://doi.org/10.3200/JACH.53.2.61-70>
- Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–196). Oxford, England: Oxford University Press.
- Tomer, J. F. (2017). *Advanced Introduction to Behavioral Economics*. Glos: UK: Edward Elgar Publishing Limited.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science (New Series)*, 185(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039–1061.

- U.S. Department of Health and Human Services. (2013). *2012 National Survey of Organ Donation Attitudes and Behaviors*. Rockville, Maryland. Retrieved from <http://organdonor.gov/dtcp/nationalsurveyorgandonation.pdf>
- Vallgård, S. (2012). Nudge a new and better way to improve health? *Health Psychology, 104*(2), 200–203. <https://doi.org/10.1016/j.healthpol.2011.10.013>
- van Dalen, H. P., & Henkens, K. (2014). Comparing the effects of defaults in organ donation systems. *Social Science & Medicine, 106*, 137–142. <https://doi.org/10.1016/j.socscimed.2014.01.052>
- Van Rooij, M. C., Lusardi, A., & Alessie, R. J. (2011). Financial literacy and retirement planning in the Netherlands. *Journal of Economic Psychology, 32*(4), 593–608.
- Vincent, A., & Logan, L. (2012). Consent for organ donation. *British Journal of Anaesthesia, 108*(suppl 1), i80–i87. <https://doi.org/10.1093/bja/aer353>
- Wakefield, M., Coomber, K., Zacher, M., Durkin, S., Brennan, E., & Scollo, M. (2015). Australian adult smokers' responses to plain packaging with larger graphic health warnings 1 year after implementation: results from a national cross-sectional tracking survey. *Tobacco Control, 24*(Suppl 2), ii17–ii25. <https://doi.org/10.1136/tobaccocontrol-2014-052050>
- Wansink, B. (2004). Environmental factors that increase the food intake and consumption volume of unknowing consumers. *Annual Review of Nutrition, 24*, 455–479. <https://doi.org/10.1146/annurev.nutr.24.012003.132140>
- Wansink, B., & van Ittersum, K. (2005). Shape of glass and amount of alcohol poured: comparative study of effect of practice and concentration. *BMJ, 331*. <https://doi.org/10.1136/bmj.331.7531.1512>
- Wansink, B., & van Ittersum, K. (2013). Portion size me: Plate-size induced consumption norms and win-win solutions for reducing food intake and waste. *Journal of Experimental Psychology: Applied, 19*(4), 320–332. <https://doi.org/10.1037/a0035053>
- Webb, K. L., Solomon, L. S., Sanders, J., Akiyama, C., & Crawford, P. B. (2011). Menu labeling responsive to consumer concerns and shows promise for changing patron purchases. *Journal of Hunger & Environmental Nutrition, 6*(2), 166–178. <https://doi.org/10.1080/19320248.2011.576210>
- Webb, O. J., & Eves, F. F. (2007). Promoting Stair Climbing: Intervention Effects Generalize to a Subsequent Stair Ascent. *American Journal of Health Promotion, 22*(2), 114–119. <https://doi.org/10.4278/0890-1171-22.2.114>

- Wells, J. (2014). *Pension Annuities: A review of consumer behaviour*. UK.
- White, M. (2013). *The Manipulation of Choice: Ethics and Libertairan Paternalism*. New York: Palgrave Macmillan.
- Whitman, D. G., & Rizzo, M. J. (2015). The Problematic Welfare Standards of Behavioural Paternalism. *Review of Philosophy and Psychology*, 6(3).
- WHO. (2015). Health Warnings. Retrieved March 3, 2017, from <http://www.euro.who.int/en/health-topics/disease-prevention/tobacco/policy/who-framework-convention-on-tobacco-control-who-fctc/key-areas-of-tobacco-control-policy/health-warnings>
- WHO. (2016). WHO | Tobacco. Retrieved March 15, 2017, from <http://www.who.int/mediacentre/factsheets/fs339/en/>
- WHO. (2017a). NCDs | Noncommunicable diseases and their risk factors. Retrieved March 2, 2017, from <http://www.who.int/ncds/introduction/en/>
- WHO. (2017b). *Plain packaging of tobacco products: evidence, design and implementation*. Retrieved from http://apps.who.int/iris/bitstream/10665/207478/1/9789241565226_eng.pdf?ua=1
- Whyte, K. P., Selinger, E., Caplan, A. L., & Sadowski, J. (2012). Nudge, Nudge or Shove, Shove—The Right Way for Nudges to Increase the Supply of Donated Cadaver Organs. *The American Journal of Bioethics*, 12(2), 32–39. <https://doi.org/10.1080/15265161.2011.634484>
- Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies*, 61, 341–355. <https://doi.org/10.1111/j.1467-9248.2012.00974.x>
- Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin*, 106(2), 231.
- Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in Everyday Life: Thought, Emotion, and Action. *Journal of Personality and Social Psychology*, 83(6), 1281–1297. <https://doi.org/10.1037//0022-3514.83.6.1281>
- World Bank. (2016). Pensions. Retrieved from <http://www.worldbank.org/en/topic/pensions>

APPENDIX B

CHAPTER 3: Underlying Wishes and Nudged Choice in Pro-self Context

Questionnaire: Experiment 1a (UK Sample)

COVER STORY

In this part of the survey, we are going to ask a number of questions on the subject of Individual Savings Account (ISA). Psychologists have found that sometimes people's true preferences (what they really want) differ from what they indicate in questionnaires. This is why we are asking you what you think a person's true preferences are in the following. Please read each question carefully and response as you see fit. There are no right or wrong answers. Again as a reminder, ISA is a voluntary, tax-free, retirement protection. The aim of ISAs was to help UK residents make money from investments, or make money from their savings. In both cases this was without having to pay tax on the money that is made. This is also why ISAs are seen as a scheme to help people prepare for retirement.

1. Out of 100 people who live in the UK, for how many do you think their true preference is to be registered in the ISA Scheme?

OPT-OUT CONDITION

Mark lives in an area in the City with an **OPT OUT** banking system. Under this system, everyone is automatically registered as ISA default, meaning that they will automatically have their money invested for them by their bank/building Society. The Bank/Building Society will decide on the individual's behalf exactly how the money will be allocated to investment schemes. Anyone who wishes NOT to be registered in the ISA default scheme MUST make an extra effort by going online and changing their preferences or by calling the bank/building society. Under the OPT OUT system Mark is REGISTERED AS ISA-Default meaning that he is entered into the ISA-DEFAULT SCHEME.

2. How likely do you think it is that Mark's true preference was to be registered in the ISA default scheme?
3. If we assembled 100 people whose true preference is to be a member of the ISA default scheme, how many of them do you think will be registered in the ISA default scheme on the OPT OUT system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the ISA default system, how many of them do you think will be registered in the ISA default scheme in the OPT OUT system?

Imagine that you live now in the same city as Mark does (in which the OPT OUT system is in place). Again, if you wish not to be a ISA-default member, you will need to make extra effort by contacting your ISA provider.

5. How likely do you think that you will be registered in the ISA default scheme? (0 - very unlikely; 100 - very likely).

MANDATED CHOICE CONDITION

Mark lives in an area in the City with a CHOICE banking system. Under this system, everyone is legally required by their bank/building society to choose between being a NON-ISA-Default MEMBER or a ISA-Default MEMEBER. Under the CHOICE system Mark is REGISTERED AS ISA-Default meaning he is entered into the ISA-DEFAULT SCHEME.

2. How likely do you think it is that Mark's true preference was to be registered in the ISA default scheme?
3. If we assembled 100 people whose true preference is to be a member of the ISA Default scheme, how many of them do you think will be registered in the ISA Default scheme on the CHOICE system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the ISA Default system, how many of them do you think will be registered in the ISA Default scheme on the CHOICE system?

Imagine that you live now in the same city as Mark does (in which the CHOICE system is in place).

5. How likely do you think that you will be registered in the ISA default scheme? (0 - very unlikely; 100 - very likely).

OPT-IN CONDITION

Mark lives in an area in the City with an OPT IN banking system. Under this system, everyone is automatically registered as NON-ISA Default, meaning that they will NOT have their money automatically invested by their bank/building society in the NON-ISA Default SCHEME. Anyone who wishes to be registered in ISA default scheme must make an extra effort by going online and changing their preferences or by calling the bank/building society. Under the OPT IN system Mark is REGISTERED AS ISA default meaning he is entered into the ISA-DEFAULT SCHEME.

2. How likely do you think it is that Mark's true preference was to be registered in ISA default scheme?
3. If we assembled 100 people whose true preference is to be a member of the ISA Default scheme, how many of them do you think will be registered in the ISA Default scheme on the OPT IN system?

4. If we assembled 100 people whose true preference is NOT TO be a member of the ISA Default system, how many of them do you think will be registered in the ISA Default scheme on the OPT-IN system?

Imagine that you live now in the same city as Mark does (in which the OPT IN system is in place). Again, if you wish to be a ISA-default member, you will need to make extra effort by contacting your ISA provider.

5. How likely do you think that you will be registered in the ISA default scheme? (0 - very unlikely; 100 - very likely)

Questionnaire: Experiment 1b (HK Sample)

In this part of the survey, we are going to ask a number of questions on the subject of Mandatory Provident Fund (MPF). Psychologists have found that sometimes people's true preferences (what they really want) differ from what they indicate in questionnaires. This is why we are asking you what you think a person's true preferences are in the following. Please read each question carefully and response as you see fit. There are no right or wrong answers. Again as a reminder, MPF is a mandatory retirement protection scheme. Almost all employers, employees and self-employed persons aged within 18 to 64 are legally required to contribute to MPF fund(s). MPF funds are offered by approved banks or insurers.

1. Out of 100 people who live in HK, for how many do you think their true preference is to be registered in the MPF Scheme?

OPT-OUT CONDITION

Mark lives in an area in the City with an **OPT OUT** banking system. Under this system, everyone is automatically registered as MPF default, meaning that they will automatically have their money invested for them by their Bank/Insurer. The Bank/Insurer will decide on the individual's behalf exactly how the money will be allocated to investment schemes. Anyone who wishes NOT to be registered in the MPF default scheme **MUST** make an extra effort by going online and changing their preferences or by calling the bank/insurer. Under the OPT OUT system Mark is **REGISTERED AS MPF-default** meaning he is entered into the **MPF-DEFAULT SCHEME**.

2. How likely do you think it is that Mark's true preference was to be registered in the MPF-default scheme?
3. If we assembled 100 people whose true preference is to be a member of the MPF default scheme, how many of them do you think will be registered in the MPF default scheme on the OPT OUT system?

4. If we assembled 100 people whose true preference is NOT TO be a member of the MPF default system, how many of them do you think will be registered in the MPF default scheme in the OPT OUT system?

Imagine that you live now in the same city as Mark does (in which the OPT OUT system is in place). Again, if you wish not to be a MPF-default member, you will need to make extra effort by contacting your MPF provider.

5. How likely do you think that you will be registered in the MPF default scheme? (0 - very unlikely; 100 - very likely).

MANDATED CHOICE CONDITION

Mark lives in an area in the City with a CHOICE banking system. Under this system, everyone is legally required by their bank/insurer to choose between being a NON-MPF-Default MEMBER or a MPF-Default MEMEBER. Under the CHOICE system Mark is REGISTERED AS MPF-Default meaning he is entered into the MPF-DEFAULT SCHEME.

2. How likely do you think it is that Mark's true preference was to be registered in the MPF default scheme?
3. If we assembled 100 people whose true preference is to be a member of the MPF Default scheme, how many of them do you think will be registered in the MPF Default scheme on the CHOICE system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the MPF Default system, how many of them do you think will be registered in the MPF Default scheme on the CHOICE system?

Imagine that you live now in the same city as Mark does (in which the CHOICE system is in place).

5. How likely do you think that you will be registered in the MPF default scheme? (0 - very unlikely; 100 - very likely).

OPT-IN CONDITION

Mark lives in an area in the City with an OPT IN banking system. Under this system, everyone is automatically registered as NON-MPF Default, meaning they will NOT have their money automatically invested by their bank/insurer in the NON-MPF DEFAULT SCHEME. Anyone who wishes to be registered in the MPF default scheme must make an extra effort by going online and changing their preferences or by calling the bank/insurer. Under the OPT IN system Mark is REGISTERED AS MPF-default meaning he is entered into the MPF-DEFAULT SCHEME.

2. How likely do you think it is that Mark's true preference was to be registered in the MPF default scheme?
3. If we assembled 100 people whose true preference is to be a member of the MPF Default scheme, how many of them do you think will be registered in the MPF Default scheme on the OPT IN system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the MPF Default system, how many of them do you think will be registered in the MPF Default scheme on the OPT-IN system?

Imagine that you live now in the same city as Mark does (in which the OPT IN system is in place). Again, if you wish to be a MPF-default member, you will need to make extra effort by contacting your MPF provider.

5. How likely do you think that you will be registered in the MPF default scheme? (0 - very unlikely; 100 - very likely).

Questionnaire: Experiment 2a (UK Sample)

OPT-OUT CONDITION

Mark lives in an area in the City with an OPT OUT banking system. Under this system, everyone is automatically registered as ISA default, meaning that they will automatically have their money invested for them by their bank/building Society. The Bank/Building Society will decide on the individual's behalf exactly how the money will be allocated to investment schemes. Anyone who wishes NOT to be registered in the ISA default scheme MUST make an extra effort by going online and changing their preferences or by calling the bank/building society. Under the OPT OUT system Mark is REGISTERED AS ISA meaning that he is entered into the ISA-DEFAULT SCHEME.

1. How likely do you think it is that Mark's true preference was to be registered in the ISA default scheme?

Recall that Mark's area in the city has an OPT OUT banking system. Under this system, everyone is automatically registered as ISA default, meaning they will automatically have their money invested for them by their bank/building society. The Bank/Building Society will decide on the individual's behalf exactly how the money will be allocated to investment schemes. Anyone who wishes NOT to be registered in the ISA default scheme must make an extra effort by going online and changing their preferences or by calling the bank/building society.

2. Out of 100 people who live in the same area as Mark, for how many do you think their true preference is to be registered in the ISA Default Scheme?

3. If we assembled 100 people whose true preference is to be a member of the ISA default scheme, how many of them do you think will be registered in the ISA default scheme on the OPT OUT system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the ISA default system, how many of them do you think will be registered in the ISA default scheme in the OPT OUT system?

MANDATED CHOICE CONDITION

Mark lives in an area in the City with a CHOICE banking system. Under this system, everyone is legally required by their bank/building society to choose between being a NON-ISA-Default MEMBER or a ISA-Default MEMEBER. Under the CHOICE system Mark is REGISTERED AS ISA-Default meaning he is entered into the ISA-DEFAULT SCHEME.

1. How likely do you think it is that Mark's true preference was to be registered in the ISA default scheme?

Recall that Mark's area in the city has a CHOICE banking system. Under this system, everyone is legally required by their bank/building society to choose between being a NON-ISA-Default MEMBER or a ISA-Default MEMEBER.

1. Out of 100 people who live in the same area as Mark, for how many do you think their true preference is to be registered in the ISA Default Scheme?
2. If we assembled 100 people whose true preference is to be a member of the ISA Default scheme, how many of them do you think will be registered in the ISA Default scheme on the CHOICE system?
3. If we assembled 100 people whose true preference is NOT TO be a member of the ISA Default system, how many of them do you think will be registered in the ISA Default scheme on the CHOICE system?

OPT-IN CONDITION

Mark lives in an area in the City with an OPT IN banking system. Under this system, everyone is automatically registered as NON-ISA Default, meaning that they will NOT have their money automatically invested by their bank/building society in the ISA SCHEME. Anyone who wishes to be registered in ISA default scheme must make an extra effort by going online and changing their preferences or by calling the bank/building society. Under the OPT IN system Mark is REGISTERED AS ISA default meaning he is entered into the ISA-DEFAULT SCHEME.

1. How likely do you think it is that Mark's true preference was to be registered in ISA default scheme?

Recall that Mark's area in the city has an OPT IN banking system. Under this system, everyone is automatically registered as NON-ISA DEFAULT MEMBER, meaning they will NOT have their money automatically invested by their bank/building society in the ISA SCHEME. Anyone who wishes to be registered in ISA default scheme must make an extra effort by going online and changing their preferences or by calling the bank/building society.

2. Out of 100 people who live in the same area as Mark, for how many do you think their true preference is to be registered in the ISA DEFAULT Scheme?
3. If we assembled 100 people whose true preference is to be a member of the ISA Default scheme, how many of them do you think will be registered in the ISA Default scheme on the OPT IN system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the ISA Default system, how many of them do you think will be registered in the ISA Default scheme on the OPT-IN system?

Questionnaire: Experiment 2b (HK Sample)

OPT-OUT CONDITION

Mark lives in an area in the City with an OPT OUT banking system. Under this system, everyone is automatically registered as ISA default, meaning that they will automatically have their money invested for them by their bank/building Society. The Bank/Building Society will decide on the individual's behalf exactly how the money will be allocated to investment schemes. Anyone who wishes NOT to be registered in the ISA default scheme MUST make an extra effort by going online and changing their preferences or by calling the bank/building society. Under the OPT OUT system Mark is REGISTERED AS ISA meaning that he is entered into the ISA-DEFAULT SCHEME.

1. How likely do you think it is that Mark's true preference was to be registered in the ISA default scheme?

Recall that Mark's area in the city has an OPT OUT banking system. Under this system, everyone is automatically registered as ISA default, meaning they will automatically have their money invested for them by their bank/building society. The Bank/Building Society will decide on the individual's behalf exactly how the money will be allocated to investment schemes. Anyone who wishes NOT to be registered in the ISA default scheme must make an

extra effort by going online and changing their preferences or by calling the bank/building society.

2. Out of 100 people who live in the same area as Mark, for how many do you think their true preference is to be registered in the ISA Default Scheme?
3. If we assembled 100 people whose true preference is to be a member of the ISA default scheme, how many of them do you think will be registered in the ISA default scheme on the OPT OUT system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the ISA default system, how many of them do you think will be registered in the ISA default scheme in the OPT OUT system?

MANDATED CHOICE CONDITION

Mark lives in an area in the City with a CHOICE banking system. Under this system, everyone is legally required by their bank/building society to choose between being a NON-ISA-Default MEMBER or a ISA- Default MEMEBER. Under the CHOICE system Mark is REGISTERED AS ISA-Default meaning he is entered into the ISA-DEFAULT SCHEME.

1. How likely do you think it is that Mark's true preference was to be registered in the ISA default scheme?

Recall that Mark's area in the city has a CHOICE banking system. Under this system, everyone is legally required by their bank/building society to choose between being a NON-ISA-Default MEMBER or a ISA-Default MEMEBER.

2. Out of 100 people who live in the same area as Mark, for how many do you think their true preference is to be registered in the ISA Default Scheme?
3. If we assembled 100 people whose true preference is to be a member of the ISA Default scheme, how many of them do you think will be registered in the ISA Default scheme on the CHOICE system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the ISA Default system, how many of them do you think will be registered in the ISA Default scheme on the CHOICE system?

OPT-IN CONDITION

Mark lives in an area in the City with an OPT IN banking system. Under this system, everyone is automatically registered as NON-ISA Default, meaning that they will NOT have their money automatically invested by their bank/building society in the ISA SCHEME. Anyone who wishes to be registered in ISA default scheme must make an extra effort by going online and changing their preferences or by calling the bank/building society. Under the OPT IN system

Mark is REGISTERED AS ISA default meaning he is entered into the ISA-DEFAULT SCHEME.

1. How likely do you think it is that Mark's true preference was to be registered in ISA default scheme?

Recall that Mark's area in the city has an OPT IN banking system. Under this system, everyone is automatically registered as NON-ISA DEFAULT MEMBER, meaning they will NOT have their money automatically invested by their bank/building society in the ISA SCHEME. Anyone who wishes to be registered in ISA default scheme must make an extra effort by going online and changing their preferences or by calling the bank/building society.

2. Out of 100 people who live in the same area as Mark, for how many do you think their true preference is to be registered in the ISA DEFAULT Scheme?
3. If we assembled 100 people whose true preference is to be a member of the ISA Default scheme, how many of them do you think will be registered in the ISA Default scheme on the OPT IN system?
4. If we assembled 100 people whose true preference is NOT TO be a member of the ISA Default system, how many of them do you think will be registered in the ISA Default scheme on the OPT-IN system?

APPENDIX C

CHAPTER 4: Underlying Wishes and Nudged Choice in Pro-social Context

Study 1 (US Sample)

COVER STORY

In this survey, we are going to ask you a number of questions on the subject of organ donation. Please read each question carefully and respond as you see fit. There are no right or wrong answers. AN ORGAN DONOR is a person who donates their organs for transplant in the event of their death.

Q1 Out of 100 people in the U.S., for how many do you think their true preference is to donate their organs?

OPT-IN CONDITION

Mark lives in an area with an OPT IN system of organ donation. Under this system, everyone is automatically registered to be a NON-DONOR, meaning they will NOT have their organs used in the event of their death. Anyone who wishes to be a donor must make an extra effort by going online and changing their preferences or by calling the donor line. Mark is involved in a fatal accident that leaves many of his vital organs intact. Under the OPT IN system Mark was REGISTERED AS AN ORGAN DONOR. We want to know how you think Mark felt about being an organ donor.

Q2 How likely do you think it is that Mark's true preference was to donate his organs?

As you know Mark is involved in a fatal accident that leaves many of his vital organs intact. John is Mark's nephew. John has just found out that his uncle Mark has unfortunately been in a fatal accident. The doctor informs John that his uncle is registered as an organ donor under the OPT IN system, but as Mark's nearest relative John has to make the final decision about what actually happens to his uncle's organs. John has no opinion one way or another about what should be done, so he wants to base his decision on his Uncle Mark's preferences.

Q3 How likely is it that John will agree to his Uncle's organs being donated?

Highly Unlikely (1)	Moderately Unlikely (2)	Likely (3)	Moderately Likely (4)	Highly Likely (5)
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4 What will John decide to do, will he donate his Uncle's organs?

- Yes
- No

Recall that Mark's area has an OPT IN system of organ donation. Under this system, everyone is automatically registered to be a NON-DONOR. This means people in the area will NOT have their organs used in the event of their death. Anyone who wishes to be a donor must make an extra effort by going online and changing their preferences or by calling the donor line.

Q5 If we assembled 100 people whose true preference is to DONATE their organs, how many of them do you think will be registered as organ donors on the OPT IN system?

Q6 If we assembled 100 people whose true preference is NOT TO DONATE their organs, how many of them do you think will be registered as organ donors on the OPT IN system?

OPT-OUT CONDITION

Mark lives in an area with an OPT OUT system of organ donation. Under this system, everyone is automatically registered to be a DONOR, meaning they will have their organs used in the event of their death. Anyone who wishes to be a NON-DONOR must make an extra effort by going online and changing their preferences or by calling the donor line. Mark is involved in a fatal accident that leaves many of his vital organs intact. Under the OPT OUT system Mark was REGISTERED AS AN ORGAN DONOR. We want to know how you think Mark felt about being an organ donor.

Q2 How likely do you think it is that Mark's true preference was to donate his organs?

As you know Mark is involved in a fatal accident that leaves many of his vital organs intact. John is Mark's nephew. John has just found out that his uncle Mark has unfortunately been in a fatal accident. The doctor informs John that his uncle is registered as an organ donor under the OPT OUT system, but as Mark's nearest relative John has to make the final decision about what actually happens to his uncle's organs. John has no opinion one way or another

about what should be done, so he wants to base his decision on his Uncle Mark's preferences.

Q3 How likely is it that John will agree to his Uncle's organs being donated?

Highly Unlikely (1)	Moderately Unlikely (2)	Likely (3)	Moderately Likely (4)	Highly Likely (5)
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4 What will John decide to do, will he donate his Uncle's organs?

- Yes
- No

Recall that Mark's area has an OPT OUT system of organ donation. Under this system, everyone is automatically registered to be a DONOR. This means that they WILL have their organs used in the event of their death. Anyone who wishes to be a NON-DONOR must make an extra effort by going online and changing their preferences or by calling the donor line.

Q5 If we assembled 100 people whose true preference is to DONATE their organs, how many of them do you think will be registered as organ donors on the OPT OUT system?

Q6 If we assembled 100 people whose true preference is NOT TO DONATE their organs, how many of them do you think will be registered as organ donors on the OPT OUT system?

MANDATED CHOICE CONDITION

Mark lives in an area with a CHOICE system of organ donation. Under this system, everyone is legally required to choose between being a DONOR OR NON-DONOR before they register for their driver's license. Mark is involved in a fatal accident that leaves many of his vital organs intact. Under the CHOICE system Mark was REGISTERED AS AN ORGAN DONOR. We want to know how you think Mark felt about being an organ donor.

Q2 How likely do you think it is that Mark's true preference was to donate his organs?

As you know Mark is involved in a fatal accident that leaves many of his vital organs intact. John is Mark's nephew. John has just found out that his uncle Mark has unfortunately been in a fatal accident. The doctor informs John that his uncle is registered as an organ donor under the CHOICE system, but as Mark's nearest relative John has to make

the final decision about what actually happens to his uncle’s organs. John has no opinion one way or another about what should be done, so he wants to base his decision on his Uncle Mark’s preferences.

Q3 How likely is it that John will agree to his Uncle’s organs being donated?

Highly Unlikely (1)	Moderately Unlikely (2)	Likely (3)	Moderately Likely (4)	Highly Likely (5)
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4 What will John decide to do, will he donate his Uncle’s organs?

- Yes
- No

Recall that Mark’s area has a CHOICE system of organ donation. Under this system, everyone is legally required to choose between being a DONOR OR NON-DONOR before they register for their driver’s license.

Q5 If we assembled 100 people whose true preference is to DONATE their organs, how many of them do you think will be registered as organ donors on the CHOICE system?

Q6 If we assembled 100 people whose true preference is NOT TO DONATE their organs, how many of them do you think will be registered as organ donors on the CHOICE system?

MADATORY DONOR CONDITION

Mark lives in an area with a DONOR system of organ donation. Under this system, everyone is automatically REGISTERED AS A DONOR. There is no option for changing this. Mark is involved in a fatal accident that leaves many of his vital organs intact. Under the DONOR system Mark was REGISTERED AS AN ORGAN DONOR. We want to know how you think Mark felt about being an organ donor.

Q2 How likely do you think it is that Mark’s true preference was to donate his organs?

As you know Mark is involved in a fatal accident that leaves many of his vital organs intact. John is Mark’s nephew. John has just found out that his uncle Mark has unfortunately been in a fatal accident. The doctor informs John that his uncle is registered as an organ donor under the DONOR system, but as Mark’s nearest relative John has to make the final decision about what actually happens to his uncle’s organs. John has no opinion one way or another

about what should be done, so he wants to base his decision on his Uncle Mark's preferences.

Q3 How likely is it that John will agree to his Uncle's organs being donated?

Highly Unlikely (1)	Moderately Unlikely (2)	Likely (3)	Moderately Likely (4)	Highly Likely (5)
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4 What will John decide to do, will he donate his Uncle's organs?

- Yes
- No

Recall that Mark's area has a DONOR system of organ donation. Under this system, everyone is automatically registered as a donor. There is no option for changing this.

Q5 If we assembled 100 people whose true preference is to DONATE their organs, how many of them do you think will be registered as organ donors on the DONOR system?

Q6 If we assembled 100 people whose true preference is NOT TO DONATE their organs, how many of them do you think will be registered as organ donors on the DONOR system?

DEMOGRAPHIC

We will now present you with the description of four different organ donation systems. OPT-IN: Under this system, everyone is automatically registered to be a NON-DONOR, meaning they will NOT have their organs used in the event of their death. Anyone who wishes to be a donor must make an extra effort by going online and changing their preferences or by calling the donor line.

OPT-OUT: Under this system, everyone is automatically registered to be a DONOR, meaning they will have their organs used in the event of their death. Anyone who wishes to be a NON-DONOR must make an extra effort by going online and changing their preferences or by calling the donor line.

CHOICE: Under this system, everyone is legally required to choose between being a DONOR OR NON-DONOR before they register for their driver's license.

MANDATORY DONOR: Under this system, everyone is automatically REGISTERED AS A DONOR. There is no option for changing this.

Q7 What system do you think the U.S. should adopt?

- Opt-In
- Opt-Out
- Choice
- Mandatory Donor

Q8 What is your gender?

- Male
- Female

Q9 What is your age?

Q10 What is your nationality?

Q11 Are you an organ donor?

- Yes
- No

Q12 What is your religion do you belong to?

Q13 Is organ donation forbidden in your religion?

- Yes
- No

Q14 Have you given blood before?

- Yes
- No

Q15 Do you know anyone who is registered as an organ donor?

- Yes
- No

Q16 Do you agree with the idea that everyone should be automatically included on the Organ Donor register with the ability to opt out if they wish?

- Yes
- No

Q17 Would you be willing to agree to donation when your loved one's wishes are unknown?

- Yes
- No

Q18 In the event of a death, which of these do you believe should apply in respect of organ donation?

- The family / close friend should have a final say on whether a deceased person's organs are donated or not.
- The deceased persons' wishes about donating their organs or not should be respected no matter what the family thinks.
- Don't know.

Q19 What is the current organ donation legislation in the U.S.?

- Opt out
- Presumed consent
- Mandatory/compulsory
- Need to carry an opt out card
- Opt in
- Family or close friend will decide
- Don't know

Study 2 (European Opt-In Country) and Study 3 (European Opt-Out Country)

COVER STORY

In this survey, we are going to ask you a number of questions on the subject of organ donation. Please read each question carefully and respond based on answers that are closest to what you think. There are no right or wrong answers. In this experiment, the country in consideration has an OPT-IN system of organ donation. Under this system, everyone is automatically registered to be a NON-DONOR, meaning they will NOT have their organs used in the event of their death. Anyone who wishes to be a DONOR must make an extra effort by going online and register their decision or by calling the donor line.

Q1 Out of 100 people living in this country, for how many do you think would want to donate their organs?

OPT-IN CONDITION

Mark currently lives in this country with an OPT-IN system of organ donation. Remember that under this system everyone is automatically registered to be a NON-DONOR, meaning they will NOT have their organs used in the event of their death. Anyone who wishes to be a DONOR must make an extra effort by going online and registering their decision or by calling the donor line. Mark is involved in a fatal accident that leaves many of his vital organs intact. Under the OPT-IN system Mark was REGISTERED AS AN ORGAN DONOR. We want to know how you think Mark felt about being an organ donor.

Q2 How likely do you think it is that Mark wanted to donate his organs?

As you know Mark is involved in a fatal accident that leaves many of his vital organs intact. John is Mark's nephew. John has just found out that his uncle Mark has unfortunately been in

a fatal accident. The doctor informs John that his uncle is registered as an organ donor under the OPT-IN system, but as Mark's nearest relative John has to make the final decision about what actually happens to his uncle's organs. John has no opinion one way or another about what should be done, so he wants to base his decision on his Uncle Mark's decision to donate.

Q3 To what extent does being registered to donate his organs under the OPT-IN system provide a clear indication that Mark wanted to donate his organs?

Q4 How likely do you think it is that John believes that Mark wanted to donate his organs?

Q5 How likely is it that John will agree to his Uncle's organs being donated?

Highly Unlikely (1)	Moderately Unlikely (2)	Likely (3)	Moderately Likely (4)	Highly Likely (5)
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Display This Question:

If How likely is it that John will agree to his Uncle's organs being donated? - Highly Unlikely Is Selected

Q6 Why do you think it is High Unlikely that John will donate Mark's organs?

- It is a highly traumatic time for relatives and it's just not something they can think about
- Lack of understanding of the organ donation process
- Denial and rejection of brain-death criteria
- The hope for a miracle
- Fear about organ trade and unknown organ destination
- Religious beliefs
- Insecurity about the brain-death diagnosis (if doctor knows you are a registered donor they won't do everything they can to save your life)
- Unsure about Mark's wish to donate
- Belief in body integrity about death
- Fear of objection by other family members
- There is a lack of evidence to indicate that Mark wanted to donate his organs
- None of the above

Display This Question:

If How likely is it that John will agree to his Uncle's organs being donated? - Moderately Unlikely Is Selected

Q6 Why do you think it is Moderately Unlikely that John will donate Mark's organs?

- It is a highly traumatic time for relatives and it's just not something they can think about
- Lack of understanding of the organ donation process
- Denial and rejection of brain-death criteria
- The hope for a miracle
- Fear about organ trade and unknown organ destination
- Religious beliefs
- Insecurity about the brain-death diagnosis (if doctor knows you are a registered donor they won't do everything they can to save your life)
- Unsure about Mark's wish to donate
- Belief in body integrity about death
- Fear of objection by other family members
- There is a lack of evidence to indicate that Mark wanted to donate his organs
- None of the above

Display This Question:

If How likely is it that John will agree to his Uncle's organs being donated? - Likely Is Selected

Q6 Why do you think it is Likely that John will donate Mark's organs?

- It is important to respect the deceased's wish
- This is a gift of life
- The act of good citizenship
- There is evidence to suggest that Mark wanted to donate his organs
- Mark actively made a choice to opt-in and donate his organs
- None of the above

Display This Question:

If How likely is it that John will agree to his Uncle's organs being donated? - Moderately Likely Is Selected

Q6 Why do you think it is Moderately Likely that John will donate Mark's organs?

- It is important to respect the deceased's wish
- This is a gift of life
- The act of good citizenship
- There is evidence to suggest that Mark wanted to donate his organs
- Mark actively made a choice to opt-in and donate his organs
- None of the above

Display This Question:

If How likely is it that John will agree to his Uncle's organs being donated? - Highly Likely Is Selected

Q6 Why do you think it is Highly Likely that John will donate Mark's organs?

- It is important to respect the deceased's wish
- This is a gift of life
- The act of good citizenship
- There is evidence to suggest that Mark wanted to donate his organs
- Mark actively made a choice to opt-in and donate his organs
- None of the above

Recall that Mark lives in a country that has an OPT-IN system of organ donation. Under this system, everyone is automatically registered to be a NON-DONOR. This means people in this country will NOT have their organs used in the event of their death. Anyone who wishes to be a DONOR must make an extra effort by going online and registering their decision or by calling the donor line.

Q7 If we assembled 100 people who want to DONATE their organs, how many of them do you think will be registered as organ donors on the OPT-IN system?

Q8 If we assembled 100 people who DO NOT WANT to DONATE their organs, how many of them do you think will be registered as organ donors on the OPT-IN system?

OPT-OUT CONDITION

In this experiment, the country in consideration has an OPT-OUT system of organ donation. Under this system, everyone is automatically registered to be a DONOR, meaning they WILL have their organs used in the event of their death. Anyone who wishes to be a NON-DONOR must make an extra effort by going online and registering their decision or by calling the donor line.

Q1 Out of 100 people living in this country, for how many do you think would want to donate their organs?

Mark currently lives in this country with an OPT-OUT system of organ donation. Remember that under this system everyone is automatically registered to be a DONOR, meaning they WILL have their organs used in the event of their death. Anyone who wishes to be a NON-DONOR must make an extra effort by going online and registering their decision or by calling the donor line. Mark is involved in a fatal accident that leaves many of his vital organs intact. Under the OPT-OUT system Mark was REGISTERED AS AN ORGAN DONOR. We want to know how you think Mark felt about being an organ donor.

Q2 How likely do you think it is that Mark wanted to donate his organs?

As you know Mark is involved in a fatal accident that leaves many of his vital organs intact. John is Mark's nephew. John has just found out that his uncle Mark has unfortunately been in a fatal accident. The doctor informs John that his uncle is registered as an organ donor under the OPT-OUT system, but as Mark's nearest relative John has to make the final decision about what actually happens to his uncle's organs. John has no opinion one way or another about what should be done, so he wants to base his decision on his Uncle Mark's decision to donate.

Q3 To what extent does being registered to donate his organs under the OPT-OUT system provide a clear indication that Mark wanted to donate his organs?

Questions 4, 5, and 6 are the same as the OPT-IN CONDITION

Recall that Mark lives in a country that has an OPT-OUT system of organ donation. Under this system, everyone is automatically registered to be a DONOR. This means people in this country WILL have their organs used in the event of their death. Anyone who wishes to be a NON-DONOR must make an extra effort by going online and registering their decision or by calling the donor line.

Q7 If we assembled 100 people who want to DONATE their organs, how many of them do you think will be registered as organ donors on the OPT-OUT system?

Q8 If we assembled 100 people who DO NOT WANT to DONATE their organs, how many of them do you think will be registered as organ donors on the OPT-OUT system?

MANDATED CHOICE CONDITION

In this experiment, the country in consideration has a CHOICE system of organ donation. Under this system, everyone is legally required to choose between being a DONOR or NON-DONOR before they register for their driver's license.

Q1 Out of 100 people living in this country, for how many do you think would want to donate their organs?

Mark currently lives in this country with a CHOICE system of organ donation. Remember that under this system, everyone is legally required to choose between being a DONOR or NON-DONOR before they register for their driver's license. Mark is involved in a fatal accident that leaves many of his vital organs intact. Under the CHOICE system Mark was REGISTERED AS AN ORGAN DONOR. We want to know how you think Mark felt about being an organ donor.

Q2 How likely do you think it is that Mark wanted to donate his organs?

As you know Mark is involved in a fatal accident that leaves many of his vital organs intact. John is Mark's nephew. John has just found out that his uncle Mark has unfortunately been in a fatal accident. The doctor informs John that his uncle is registered as an organ donor under the CHOICE system, but as Mark's nearest relative John has to make the final decision about

what actually happens to his uncle's organs. John has no opinion one way or another about what should be done, so he wants to base his decision on his Uncle Mark's decision to donate.

Q3 To what extent does being registered to donate his organs under the CHOICE system provide a clear indication that Mark wanted to donate his organs?

Questions 4, 5, and 6 are the same as the OPT-IN CONDITION

Recall that Mark lives in a country that has a CHOICE system of organ donation. Under this system, everyone is legally required to choose between being a DONOR or NON-DONOR before they register for their driver's license.

Q7 If we assembled 100 people who want to DONATE their organs, how many of them do you think will be registered as organ donors on the CHOICE system?

Q8 If we assembled 100 people who DO NOT WANT to DONATE their organs, how many of them do you think will be registered as organ donors on the CHOICE system?

MANDATORY DONOR CONDITION

In this experiment, the country in consideration has a DONOR system of organ donation. Under this system, everyone is automatically REGISTERED AS A DONOR. There is no option for changing this.

Q1 Out of 100 people living in this country, for how many do you think would want to donate their organs?

Mark currently lives in this country with a DONOR system of organ donation. Remember that under this system, everyone is automatically REGISTERED AS A DONOR. There is no option for changing this. There is no option for changing this. Mark is involved in a fatal accident that leaves many of his vital organs intact. Under the DONOR system Mark was REGISTERED AS AN ORGAN DONOR. We want to know how you think Mark felt about being an organ donor.

Q2 How likely do you think it is that Mark wanted to donate his organs?

As you know Mark is involved in a fatal accident that leaves many of his vital organs intact. John is Mark's nephew. John has just found out that his uncle Mark has unfortunately been in a fatal accident. The doctor informs John that his uncle is registered as an organ donor under the DONOR system, but as Mark's nearest relative John has to make the final decision about

what actually happens to his uncle's organs. John has no opinion one way or another about what should be done, so he wants to base his decision on his Uncle Mark's decision to donate.

Q3 To what extent does being registered to donate his organs under the DONOR system provide a clear indication that Mark wanted to donate his organs?

Questions 4, 5, and 6 are the same as the OPT-IN CONDITION

Recall that Mark lives in a country that has a DONOR system of organ donation. Under this system, everyone is automatically REGISTERED AS A DONOR. There is no option for changing this.

Q7 If we assembled 100 people who want to DONATE their organs, how many of them do you think will be registered as organ donors on the DONOR system?

Q8 If we assembled 100 people who DO NOT WANT to DONATE their organs, how many of them do you think will be registered as organ donors on the DONOR system?

DEMOGRAPHIC

We will now present you with the description of four different organ donation systems:

OPT-IN: Under this system, everyone is automatically registered to be a NON-DONOR, meaning they will NOT have their organs used in the event of their death. Anyone who wishes to be a DONOR must make an extra effort by going online and registering their decision or by calling the donor line.

OPT-OUT: Under this system, everyone is automatically registered to be a DONOR, meaning they will have their organs used in the event of their death. Anyone who wishes to be a NON-DONOR must take extra effort by going online and registering their decision or by calling the donor line.

CHOICE: Under this system, everyone is legally required to choose between being a DONOR OR NON-DONOR before they register for their driver's license.

MANDATORY DONOR: Under this system, everyone is automatically REGISTERED AS A DONOR. There is no option for changing this.

Q9 What is your nationality?

Q10 What system do you think your country should adopt?

- Opt-In
- Opt-Out
- Choice
- Mandatory Donor

Q11 Gender

- Male
- Female

Q12 Age

Q13 Are you an organ donor?

- Yes
- No

Q14 What religion do you belong to?

Q15 Is organ donation forbidden in your religion?

- Yes
- No

Q16 Have you given blood before?

- Yes
- No

Q17 Do you know anyone who is registered as an organ donor?

- Yes
- No

Q18 Do you agree with the idea that everyone should be automatically included on the Organ Donor Register with the ability to opt-out if they wish?

- Yes
- No

Q19 Would you be willing to agree to donation when your loved one's wishes are unknown?

- Yes
- No

Q20 In the event of a death, which of these do you believe should apply in respect of organ donation?

- The family/close friend should have a final say on whether a deceased person's organs are donated or not
- The deceased person's wishes about donating their organs or not should be respected no matter what the family thinks
- Don't know

Q21 What is the current organ donation legislation in your country?

- Opt out/ Presumed consent
- Mandatory/compulsory
- Need to carry a donor card
- Opt in
- Family or close friend will decide
- Don't know

APPENDIX D

CHAPTER 6: Transparency on Autonomy Grounds

COVER STORY

Behaviour plays an important role in people's health and wellbeing. There's evidence to show that these behaviours are deeply embedded in people's social, economic and environmental context. Recently, **psychological methods** have been used in many day-to-day situations (canteens, supermarkets, doctors waiting rooms, leisure centres, high street banks) to bring about behavior change (e.g., quit smoking, eat healthily, reduce alcohol consumption, increase physical activity, etc.). Typical psychological methods involve changing the type of information we use to base our decisions. This can be done by presenting the order of information in certain ways, highlighting certain information to make it more distinctive, and even simplifying the information itself to get the message across more directly. All of these methods are designed to help guide people to make the best decision for their own health and wellbeing.

Behaviour Agent (Participants will be presented with one of the description below)

The **Government in this country** is using psychological research to help develop a set of simple methods that adjust the way information is presented, so that it can help people to make better decisions. The reason for using psychological methods is to help improve people's behavior, because in many day-to-day contexts people may not make a decision that is best for their own health, wellbeing, and their happiness.

The **Top Advertising Company in the country** is using psychological research to help develop a set of simple methods that adjust the way information is presented, so that it can help people to make better decisions. The reason for using psychological methods is to help improve people's behavior, because in many day-to-day contexts people may not make a decision that is best for their own health, wellbeing, and their happiness.

The **Top Researchers in laboratories across this country** are using psychological research to help develop a set of simple methods that adjust the way information is presented, so that it can help people to make better decisions. The reason for using psychological methods is to

help improve people's behavior, because in many day-to-day contexts people may not make a decision that is best for their own health, wellbeing, and their happiness.

Experiment 1 (Positive Rationale for Interventions)

Transparent Behaviour Context

[SMOKING]

Recommended Psychological Method: Design cigarette packaging so that it incorporates graphic pictures of damaged lungs and warnings such as 'Smoking seriously harms you and others around you', "Smoking harms your unborn baby'.

Argument for it to work: By highlighting the negative physical and moral issues concerning smoking, the negative experiences will become more obviously associated with smoking, and this will encourage smokers to reduce or even stop smoking.

[FOOD & NUTRITION]

Recommended Psychological Method: Design packaging on food so that the front label includes nutritional information, by using a simple traffic light system (red, amber, green) to indicate how much saturated fat, salt and sugar, and calories are in food products.

Argument for it to work: By making it easier for people to interpret the nutritional content of food items through a traffic-light labelling system, people will be more aware of which foods are healthier than others, and in turn adopt/maintain a healthier diet.

[ALCOHOL]

Recommended Psychological Method: Design signage in pubs and restaurants so that they include messages such as the following: "men and women are advised not to regularly drink more than 14 units a week" and "spread your drinking over three days or more if you drink as much as 14 units a week".

Argument for it to work: By informing people about the actual appropriate amount of alcohol consumption that is reasonable to be consumed in a typical week, because people will be more aware of exceeding the limit and this should in turn reduce alcohol overconsumption.

[EXERCISE]

Recommended Psychological Method: Design stairwells with 'point-of-choice' signage that displays messages about the health advantages of taking the stairs, such as 'Stair climbing

burns more calories per minute than tennis’, ‘7 minutes of stair climbing per day protects your heart’, etc.

Argument for it to work: By presenting messages at strategic positions, people will be encouraged to use stairs instead of lifts or escalators/elevators, and this in turn will encourage people to value being more active and in turn, exercise more in general.

[FINANCE]

Recommended Psychological Method: Design investment schemes in such a way so that customers can evaluate the associated riskiness of each product based on a traffic light system; red indicates highly risky, green indicates low risk.

Argument for it to work: By making it easier for people to interpret the riskiness of an investment scheme through a traffic-light labelling system, people will be more aware of which financial products are riskier than others, and in turn help them make a better financial decision.

Non-Transparent Behaviour Context

[SMOKING]

Recommended Psychological Method: Increasing the length of the filter by 10mm and at the same time reduce the length of the cigarette to 60mm.

Argument for it to work: If a standard cigarette is 70mm then cutting back on the harmful chemicals and replacing it with more filter would seem as if the size of cigarette haven’t changed but the amount of harmful chemical is reduced. By reducing nicotine content adequately, this method helps smokers gradually adapt to lower nicotine levels, and this will encourage smokers to reduce or even stop smoking.

[FOOD & NUTRITION]

Recommended Psychological Method: Design the size of plates so that the quantity of food on them is adjusted. Large plates and bowls can make servings of food appear smaller, whereas smaller plates can lead people to misjudge that very same quantity of food as being significantly larger.

Argument for it to work: By making the plates smaller, people would be better able to adjust the amount of food they put on their plate, and avoid overconsumption of food, and in turn adopt/maintain a healthier diet.

[ALCOHOL]

Recommended Psychological Method: Design the glassware used in pubs and restaurants in such a way so that straight glasses are used, because relative to curvy glasses, it is easier to judge and pace the amount of alcohol consumed.

Argument for it to work: By changing the containers that are used to serve alcohol, this will in turn reduce the actual amount of alcohol consumed at any one sitting, and this should in turn reduce alcohol overconsumption.

[EXERCISE]

Recommended Psychological Method: Design stairwells by hanging artworks. Pictures are changed periodically to keep stair users to prolong effectiveness.

Argument for it to work: By presenting artworks along the stairwell this is more likely to encourage people to use stairs instead of lifts or escalators/elevators, and this in turn will encourage people to value being more active and in turn, exercise more in general.

[FINANCE]

Recommended Psychological Method: Design investment schemes with an automatic enrolment system so Bank/Building Society will decide on an individual's behalf exactly how the money will be allocated to investment schemes. Although if the individual didn't want it, they could opt-out of the scheme, this would involve filling in relevant paperwork.

Argument for it to work: Because people find it difficult to think about their future financial status, so making the investment schemes a default would encourage people to invest their savings, and in turn help them make a better financial decision.

Experiment 2 (Positive & Negative Rationale for Interventions)

*Experiment adopted the same experimental set up as Experiment 1, except with the additional presentation of negative rationale for interventions.

Transparent Behaviour Context

[SMOKING]

Recommended Psychological Method: Design cigarette packaging so that it incorporates graphic pictures of damaged lungs and warnings such as 'Smoking seriously harms you and others around you', 'Smoking harms your unborn baby'.

Argument for it to work: By highlighting the negative physical and moral issues concerning smoking, the negative experiences will become more obviously associated with smoking, and this will encourage smokers to reduce or even stop smoking.

Argument for it NOT to work: By highlighting the negative physical and moral issues concerning smoking, smokers will feel more defensive of their smoking habit, and as a result, smokers will end up smoking more, meaning that the method will lead to increases in smoking.

[FOOD & NUTRITION]

Recommended Psychological Method: Design *packaging* on food so that the front label includes nutritional information, by using a simple traffic light system (red, amber, green) to indicate how much saturated fat, salt and sugar, and calories are in food products.

Argument for it to work: By making it easier for people to interpret the nutritional content of food items through a traffic-light labelling system, people will be more aware of which foods are healthier than others, and in turn adopt/maintain a healthier diet.

Argument for it NOT to work: By making it easier for people to interpret the nutritional content of food items, people change their eating habit and as a result consume more food to compensate for eating healthily, meaning that this method increases people's overall daily calorie intake.

[ALCOHOL]

Recommended Psychological Method: Design signage in pubs and restaurants so that they include messages such as the following: "men and women are advised not to regularly drink more than 14 units a week" and "spread your drinking over three days or more if you drink as much as 14 units a week".

Argument for it to work: By informing people about the actual appropriate amount of alcohol consumption that is reasonable to be consumed in a typical week, because people will be more aware of exceeding the limit and this should in turn reduce alcohol overconsumption.

Argument for it NOT to work: By informing people about the actual appropriate amount of alcohol consumption that is reasonable to be consumed in a typical week, those who drink

lightly will consume more alcohol as they believe it is safe to drink 14 units a week, meaning that this method will increase overall alcohol consumption.

[EXERCISE]

Recommended Psychological Method: Design stairwells with ‘point-of-choice’ signage that displays messages about the health advantages of taking the stairs, such as ‘Stair climbing burns more calories per minute than tennis’, ‘7 minutes of stair climbing per day protects your heart’, etc.

Argument for it to work: By presenting messages at strategic positions, people will be encouraged to use stairs instead of lifts or escalators/elevators, and this in turn will encourage people to value being more active and in turn, exercise more in general.

Argument for it NOT to work: By presenting messages at strategic positions, people will avoid the stairs and hence the messages for the reason that they don’t want to feel guilty about not exercising enough, meaning that this method will lead people to being less active.

[FINANCE]

Recommended Psychological Method: Design investment schemes in such a way so that customers can evaluate the associated riskiness of each product based on a traffic light system; red indicates highly risky, green indicates low risk.

Argument for it to work: By making it easier for people to interpret the riskiness of an investment scheme through a traffic-light labelling system, people will be more aware of which financial products are riskier than others, and in turn help them make a better financial decision.

Argument for it NOT to work: By making it easier for people to interpret the riskiness of an investment scheme through the traffic-light labelling system, the method highlights the potential financial gains through risky choices, meaning that it will lead people to taking more gambles with their money and be worse off in the long run.

Non-Transparent Behaviour Context

[SMOKING]

Recommended Psychological Method: Increasing the length of the filter by 10mm and at the same time reduce the length of the cigarette to 60mm.

Argument for it to work: If a standard cigarette is 70mm then cutting back on the harmful chemicals and replacing it with more filter would seem as if the size of cigarette haven't changed but the amount of harmful chemical is reduced. By reducing nicotine content adequately, this method helps smokers gradually adapt to lower nicotine levels, and this will encourage smokers to reduce or even stop smoking.

Argument for it NOT to work: By reducing the length of the cigarette and cutting back on nicotine content, people will change their smoking habit and as a result smoke more to compensate for the shorter cigarette, meaning that this method will lead to people increasing their consumption of cigarettes.

[FOOD & NUTRITION]

Recommended Psychological Method: Design the size of plates so that the quantity of food on them is adjusted. Large plates and bowls can make servings of food appear smaller, whereas smaller plates can lead people to misjudge that very same quantity of food as being significantly larger.

Argument for it to work: By making the plates smaller, people would be better able to adjust the amount of food they put on their plate, and avoid overconsumption of food, and in turn adopt/maintain a healthier diet.

Argument for it NOT to work: By making the plates smaller, people change their eating pattern and as a result consume more food to compensate for the smaller plate, meaning that this method will lead to increases in peoples overall daily calorie intake.

[ALCOHOL]

Recommended Psychological Method: Design the glassware used in pubs and restaurants in such a way so that straight glasses are used, because relative to curvy glasses, it is easier to judge and pace the amount of alcohol consumed.

Argument for it to work: By changing the containers that are used to serve alcohol, this will in turn reduce the actual amount of alcohol consumed at any one sitting, and this should in turn reduce alcohol overconsumption.

Argument for it NOT to work: By changing the shape of the containers that are used to serve alcohol, people change their drinking habit and as a result consume more alcohol to compensate for the smaller container, meaning that this method will increase people's overall alcohol consumption.

[EXERCISE]

Recommended Psychological Method: Design stairwells by hanging artworks. Pictures are changed periodically to keep stair users to prolong effectiveness.

Argument for it to work: By presenting artworks along the stairwell this is more likely to encourage people to use stairs instead of lifts or escalators/elevators, and this in turn will encourage people to value being more active and in turn, exercise more in general.

Argument for it NOT to work: The artwork along the stairwell is not changed regularly enough, and people get bored looking at it, and so to avoid looking at it most people end up taking the lift to avoid it, and in turn overall get less exercise.

[FINANCE]

Recommended Psychological Method: Design investment schemes with an automatic enrolment system so Bank/Building Society will decide on an individual's behalf exactly how the money will be allocated to investment schemes. Although if the individual didn't want it, they could opt-out of the scheme, this would involve filling in relevant paperwork.

Argument for it to work: Because people find it difficult to think about their future financial status, so making the investment schemes a default would encourage people to invest their savings, and in turn help them make a better financial decision.

Argument for it NOT to work: The default investment scheme doesn't take into account the fact that people have different needs because their lifestyles are different, and as a result the scheme means that in the long run, overall people will end up saving less.

Experiment 3 (Negative Rationale for Interventions)

*Experiment adopted the same experimental set up as Experiment 1, exception negative rationale for interventions are presented instead of positive rationale.

Transparent Behaviour Context

[SMOKING]

Recommended Psychological Method: Design cigarette packaging so that it incorporates graphic pictures of damaged lungs and warnings such as 'Smoking seriously harms you and others around you', 'Smoking harms your unborn baby'.

Argument for it NOT to work: By highlighting the negative physical and moral issues

concerning smoking, smokers will feel more defensive of their smoking habit, and as a result, smokers will end up smoking more, meaning that the method will lead to increases in smoking.

[FOOD & NUTRITION]

Recommended Psychological Method: Design *packaging* on food so that the front label includes nutritional information, by using a simple traffic light system (red, amber, green) to indicate how much saturated fat, salt and sugar, and calories are in food products.

Argument for it NOT to work: By making it easier for people to interpret the nutritional content of food items, people change their eating habit and as a result consume more food to compensate for eating healthily, meaning that this method increases people's overall daily calorie intake.

[ALCOHOL]

Recommended Psychological Method: Design signage in pubs and restaurants so that they include messages such as the following: "men and women are advised not to regularly drink more than 14 units a week" and "spread your drinking over three days or more if you drink as much as 14 units a week".

Argument for it NOT to work: By informing people about the actual appropriate amount of alcohol consumption that is reasonable to be consumed in a typical week, those who drink lightly will consume more alcohol as they believe it is safe to drink 14 units a week, meaning that this method will increase overall alcohol consumption.

[EXERCISE]

Recommended Psychological Method: Design stairwells with 'point-of-choice' signage that displays messages about the health advantages of taking the stairs, such as 'Stair climbing burns more calories per minute than tennis', '7 minutes of stair climbing per day protects your heart', etc.

Argument for it NOT to work: By presenting messages at strategic positions, people will

avoid the stairs and hence the messages for the reason that they don't want to feel guilty about not exercising enough, meaning that this method will lead people to being less active.

[FINANCE]

Recommended Psychological Method: Design investment schemes in such a way so that customers can evaluate the associated riskiness of each product based on a traffic light system; red indicates highly risky, green indicates low risk.

Argument for it NOT to work: By making it easier for people to interpret the riskiness of an investment scheme through the traffic-light labelling system, the method highlights the potential financial gains through risky choices, meaning that it will lead people to taking more gambles with their money and be worse off in the long run.

Non-Transparent Behaviour Context

[SMOKING]

Recommended Psychological Method: Increasing the length of the filter by 10mm and at the same time reduce the length of the cigarette to 60mm.

Argument for it NOT to work: By reducing the length of the cigarette and cutting back on nicotine content, people will change their smoking habit and as a result smoke more to compensate for the shorter cigarette, meaning that this method will lead to people increasing their consumption of cigarettes.

[FOOD & NUTRITION]

Recommended Psychological Method: Design the size of plates so that the quantity of food on them is adjusted. Large plates and bowls can make servings of food appear smaller, whereas smaller plates can lead people to misjudge that very same quantity of food as being significantly larger.

Argument for it NOT to work: By making the plates smaller, people change their eating pattern and as a result consume more food to compensate for the smaller plate, meaning that this method will lead to increases in peoples overall daily calorie intake.

[ALCOHOL]

Recommended Psychological Method: Design the glassware used in pubs and restaurants in such a way so that straight glasses are used, because relative to curvy glasses, it is easier to

judge and pace the amount of alcohol consumed.

Argument for it NOT to work: By changing the shape of the containers that are used to serve alcohol, people change their drinking habit and as a result consume more alcohol to compensate for the smaller container, meaning that this method will increase people's overall alcohol consumption.

[EXERCISE]

Recommended Psychological Method: Design stairwells by hanging artworks. Pictures are changed periodically to keep stair users to prolong effectiveness.

Argument for it NOT to work: The artwork along the stairwell is not changed regularly enough, and people get bored looking at it, and so to avoid looking at it most people end up taking the lift to avoid it, and in turn overall get less exercise.

[FINANCE]

Recommended Psychological Method: Design investment schemes with an automatic enrolment system so Bank/Building Society will decide on an individual's behalf exactly how the money will be allocated to investment schemes. Although if the individual didn't want it, they could opt-out of the scheme, this would involve filling in relevant paperwork.

Argument for it NOT to work: The default investment scheme doesn't take into account the fact that people have different needs because their lifestyles are different, and as a result the scheme means that in the long run, overall people will end up saving less.

Definition Transparency (This is presented after participants read about each context)

There are two types of psychological methods: **Transparent** and **non-transparent**.

A **transparent** psychological method works in such a way that **anyone can easily identify** the actual psychological method used to change their behaviour, as well as **easily identify** how their behavior is changed by it.

A **non-transparent** psychological method works in such a way that **no one can identify** the actual psychological method used to change their behaviour, and **no one can identify** how their behavior is changed by it.

1. To what extent is it easy for you to identify the actual psychological method used to change your behaviour? Please indicate your answer on the slider between [I cannot

- easily identify the psychological method used to change my behaviour0– I can easily identify the psychological method used to change my behaviour100]
2. To what extent is it easy for you to identify **HOW** your behaviour is going to be changed by the psychological method? Please indicate your answer on the slider between [I cannot easily identify how my behaviour is changed by the psychological method0– I can easily identify how my behaviour is changed by the psychological method100]
 3. Is the psychological method described above transparent or non-transparent?
 - Transparent
 - Non-transparent
 4. To what extent do you want to change your behaviour through the psychological method in this particular situation? Please indicate your answer on a scale between [Not at all1 / Very Much9]
 5. To what extent do you think the psychological method described above would positively change **YOUR** behaviour?
Please indicate your answer on a scale between [Much less likely1 / Much more likely9]
 6. To what extent do you think it is acceptable to use the psychological method described in this context to change your behaviour?

Demographic Questions

1. Age
2. Gender
 - Male
 - Female
3. Education Level
 - High School
 - Diploma/foundation
 - Bachelors degree
 - Masters' degree
 - PhD
 - Others
4. Political Affiliation
 - Left
 - Centre
 - Right
 - Other
5. Religion
 - Religious
 - Unsure
 - Not Religious