

Behavioral Policies, Evidence, and Expertise

Gedragsbeleid, bewijs en expertise

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

donderdag 8 juni om 10:30 uur

door

Daphne Clementine Truijens
geboren te Dordrecht

Promotiecommissie:

Promotoren: Dr. H.C.K. Heilmann
Prof.dr. J.J. Vromen

Overige leden: Prof.dr. W.B.F. Brouwer
Prof.dr. L.H.J. Noordegraaf - Eelens
Prof.dr. E.M. Sent

Cover design: Bart Bergmeijer

ISBN: 978-90-833330-0-7

Printed by: Print Service Ede - Ede, The Netherlands

© Daphne Clementine Truijens, 2023

All rights reserved. No part of this publication may be reproduced, stored in retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission in writing from the proprietor(s).

Contents

1. Introduction	57
1.1 <i>Introduction – broad context</i>	57
1.2 <i>A brief history of rationality & decision-making in economics</i>	59
1.2.1 Expected Utility Theories	59
1.2.2 Bounded Rationality	61
1.2.3 Behavioral economics & the Heuristics and Biases Program	62
1.2.3.1 Kahneman and Tversky.....	62
1.2.3.2 Prospect Theory	63
1.2.4 Ecological Rationality & the Simple Heuristics Program	64
1.2.5 From theoretical programs to behavioral policies	65
1.3 <i>Evidence-Based Behavioral Policies</i>	66
1.3.1 Evidence-based behavioral policies	66
1.3.2 Nudge, Boost and combinations	66
1.3.3 Critiques on behavioral policies as evidence-based policies	68
1.3.4 Values in science.....	69
1.4 <i>Preview of the chapters</i>	71
References.....	75
Part 1. BEHAVIOR CHANGE	88
2. Behavioral policies, mechanisms, and the individuation challenge	37
<i>Abstract</i>	37
2.1 <i>Introduction</i>	37
2.2 <i>The mechanistic evidence challenge for behavioral policies</i>	39
2.3 <i>Surgery: a framing experiment</i>	40
2.3.1 Introducing “Surgery”	40
2.3.2 Behavioral Economics interpretation of Surgery	41
2.3.3 Nudge and “Surgery”	41
2.3.4 Alternative interpretations	42
2.4 <i>Behavioral economics’ grounds for the selection of options</i>	43
2.5 <i>The individuation challenge for behavioral policies</i>	45
2.5.1 Cognitive criticisms	45
2.5.2 Normative criticism.....	45
2.5.3 The individuation challenge	47
2.6 <i>A different mechanistic evidence challenge?</i>	48
2.7 <i>Conclusion</i>	49
References.....	50
3. Views on deceased organ donation in the Netherlands	92
<i>Abstract</i>	92
3.1 <i>Introduction</i>	92
3.2 <i>Methods</i>	93
3.2.1 Q-methodology.....	93
3.2.2 Development of the statement set	94

3.2.3 Data collection	95
3.2.4. Data analysis.....	97
3.3 Results.....	97
Viewpoint 1 – Not donating my organs would be a waste	99
Viewpoint 2 – It does not go with my religion.....	99
Viewpoint 3 – My family should decide.....	100
Viewpoint 4 - It is a good deed, but I’m doubtful	101
3.4 Discussion	102
References.....	106
Part 2. EVIDENCE-BASED BEHAVIORAL POLICY.....	111
4. Coherence between theory and policy in Nudge and Boost.....	114
<i>Abstract</i>	114
4.1 Introduction.....	114
4.2 <i>A first look on theory-policy coherence in Nudge and Boost</i>	116
4.3 <i>Theory-policy coherence by Grüne-Yanoff and Hertwig</i>	119
4.3.1 A more sophisticated view on theory-policy coherence	119
4.3.2 Elements for Evaluation of Theory-Policy Coherence	119
4.3.3 Coherence.....	121
4.3.4 Challenges for Theory-Policy Coherence	122
4.3.4.1 Coherence as consistency.....	122
4.3.4.2 Beyond necessary policy assumptions	123
4.4. <i>The relevance of theory-policy coherence for evidence-based policy</i>	124
4.4.1 The choice for the policymaker	124
4.4.2 What theory-policy coherence can do for the evidence-based policymaker.....	126
4.4.2.1 Two challenges in evidence-based behavioral policy	126
4.4.2.2 Assessing (ir)relevant (mechanistic) evidence	127
4.4.2.3 Guide to what kind of mechanistic evidence is needed.....	127
4.4.2.4 Assessing seemingly conflicting evidence.....	127
4.4.2.5 ‘Crisscross coherence’.....	128
4.4.2.6 Other theories and policies	130
4.5 <i>Conclusions</i>	131
References.....	132
5. Behavioral policies, evidence and expertise.....	107
<i>Abstract</i>	107
5.1 <i>Evidence for Behavioral Policies</i>	107
5.2 <i>Towards a nuanced view on the evidence-base for behavioral policies</i>	109
5.3 <i>Cognitive mechanisms</i>	110
5.3.1 Two competing research traditions	111
5.3.2 Implications of different assumptions	112
5.3.2.1 Extrapolation.....	112
5.3.2.2 Long term effects	112
5.3.2.3 Preference elicitation.....	112
5.4 <i>Rationality</i>	113
5.4.1 Different theories of rationality	113
5.4.2 Implications for the policymaker.....	114
5.5 <i>How to engage with behavioral experts and expertise?</i>	115
5.5.1 Engaging expertise.....	115

5.5.2 3xE Framework- Tools for expert and expertise engagement	116
5.5.2.1 Expert panels	117
5.5.2.2 Evidence-gap-maps & systematic literature reviews	117
5.5.2.3 Evaluating (lists of) Alternatives.....	117
5.6 <i>Conclusion</i>	118
<i>References</i>	119
6. Conclusions	140
6.1 <i>Back to where it began</i>	140
6.2 <i>Evidential challenges for behavioral policies</i>	140
6.3 <i>Implications for evidence-based policy</i>	143
Summary	146
Nederlandse samenvatting	148
Appendix 1: Academic CV	136
<i>Education</i>	136
<i>Employment</i>	136
<i>Boards and Committees</i>	138
<i>Academic Publications</i>	138
<i>Academic Talks</i>	138
<i>Supervised Publications for Public Policy</i>	139
<i>Public (Policy) Talks and Interviews</i>	139

Chapter 1.

Introduction

1. Introduction

1.1 Introduction – broad context

Behavior change has always been high on the agenda of governments, business, educators, economists, and individuals themselves. For instance, even though many people seem to have positive attitude towards organ donation, organ donor registrations are lagging, and shortages are increasing in many countries. Furthermore, people often value their health and realize they need to stick to nutritious food in proportionate sizes. Yet even if people want to commit to certain diets, they often fail to do so. Meanwhile obesity and other diet-related health problems have become a matter of public health in most Western countries. Problems like these are widespread and solutions have traditionally been sought in informational campaigns, financial disincentives, and other policies that aim to either update people's beliefs by providing more information or implementing legal or economic incentives. However, recent discoveries in psychology and behavioral economics show that small changes in how choice options are presented can have large effects on people's behavior. For instance, opt-out systems lead to a significant increase of organ donor registrations (e.g., Abadie and Gay 2006, Dinner et al. 2011, Johnson and Goldstein 2003, 2004, Shepherd, O'Carroll and Ferguson 2014), and even simple adjustments such as smaller plates and the placement of healthy food options at eye level often lead to healthier food choices (Ensaff et al. 2015, Van Gestel, Kroese and de Ridder 2018, Cadarion and Chandon 2020, Mikkelsen et al. 2021). These findings show that individuals often diverge from the standards of rationality set by mainstream economics, such as rules of first order logic and Bayesian probability, and criteria for internal consistency of preferences such as completeness, transitivity, and stability. Therefore, it seems that people are often irrational decision-makers, or at the very least make sub-optimal decisions some of the times.

The so-called "behavioral insights" from behavioral economics and psychology have inspired researchers and policymakers alike to develop interventions, which help people make better decisions. It even catalyzed an "evidence-based policy movement" in which many governments have implemented behavioral insight units or teams (Halpern 2014, OECD 2017). In these units, policy makers draw on scientific expertise about behavior to tackle complex societal issues such as energy usage, health decisions, and organ donor shortages. Behavioral policies use the small changes in the choice environment such as defaults to bring about "better" individual or societal outcomes (e.g., Thaler and Sunstein 2008, Rebonato 2012, Hansen 2016). Quintessential about behavioral policies is that they are motivated by experimental insights. Here, it is critical that these are subject to interpretation by the researchers. Behavioral policies are motivated by the idea that the research shows that people are often irrational decision-makers and that these policies can rectify the negative outcomes of these irrationalities (e.g., Thaler and Sunstein 2008, Hansen 2016, Sunstein 2017, Mongin and Cozic 2018). For instance, to increase organ donor registrations, several countries have made organ donor registration the default (e.g., Shepherd, O'Carroll and Ferguson 2014). That is, these countries have switched from opt-in systems to opt-out systems. In opt-out systems citizens are automatically registered as (deceased) organ donors unless they actively unregister. Advocates of behavioral policies generally argue that this triggers one or several cognitive processes in people. For instance, one of the explanations of the default effect is that people lack willpower (e.g., Johnson and Goldstein 2003, Johnson and Goldstein 2004, Abadie and

Gay 2006, Thaler and Sunstein 2008, Kaiser, Arnold and Otto 2014). People procrastinate or refrain altogether from signing up as organ donors, even though they show positive attitudes towards registering (e.g., Bernhardt and Reichenspurner 2014, Marck et al. 2014, Marlais et al. 2017, Wall et al. 2015). Defaults then, tap into people's decision-making processes and influence people to choose the option that is the status quo.

However, the evidence for how and why people really show behavior change is often controversial (e.g., Smith, Goldstein and Johnson 2013, Gigerenzer 2015, Jachimowicz, Duncan and Weber 2016, Wilson et al. 2016, Grüne-Yanoff 2016). There are many different experiments, with different interpretations, and different evaluations. It is unclear how people view the choice options and why they choose the way that they do. As a consequence, behavioral policies may not have the anticipated effect. For instance, even though defaults are often advocated as one of the most successful behavioral policies, opt out systems have backfired in some cases. In some countries, citizens understood the changed default for organ donation as a 'trick' from the government. Rather than seeing the default option as one that helps register without much willpower and cognitive effort, they saw it as a form of manipulation. This led to, at least temporarily, a reduction of donor registrations instead of an increase (e.g., Krijnen, Tannenbaum and Fox 2017 about the Netherlands, Jachimowicz et al. 2019 more generally). Furthermore, even if defaults do lead to more registrations, this does not necessarily imply that also more organs will indeed be donated. People may stick with defaults but let family members make the final decision (e.g., Delgado et al. 2019, Truijens and van Exel 2019). In both cases, people have different views about the decision problem than researchers assumed. Furthermore, these views are not necessarily irrational. Sticking with the default because ultimately the family will decide does save time and energy. In the end, people are indifferent about the default. So, while there is a lot of experimental evidence for behavioral policies, the interpretation of this evidence is complex. First, it is unclear how people view the decision problem, which is crucial for whether the policy will be successful or not. Second, without this, policies cannot be justified based on the idea that people are often irrational makers and that the policies help make better decisions. Hence, interpreting the existing experimental evidence in the right way is crucial to develop efficient and effective behavioral policies.

This thesis aims to provide insight in how both the empirical evidence for behavioral policies and normative aspects of their interpretation are generated and what this implies for evidence-based policymaking and policymakers. The driving question is: on which underlying normative and empirical assumptions is the interpretation of evidence for behavioral policies based? More specifically, what is the role of agents' perspectives in this? I argue that the interpretation of the experimental evidence underlying behavioral policies often draws on cognitive theories about people's perspectives on the choice options. However, these views are presupposed rather than firmly established by the evidence, or so I will show. Consequently, the policymaker cannot justify the behavioral policy design and its proposed welfare benefits as evidence-based without further argument. In light of this, the thesis advocates for a more prominent role of people's own perspectives in the research analysis for behavioral policies. Furthermore, the evidence-base of behavioral policies is also controversial in a more general sense. Many empirical and normative assumptions are highly contested amongst researchers working from different research programs and traditions. In addition, even the connection between the policies and the theoretical research programs from which they originated is often weak. The thesis therefore suggests a more thorough

evaluation of the connection between the behavioral theory and the policy assumptions. Even though evaluating the evidence-base of behavioral policies based on “theory-policy coherence” comes with its own limitations, it does assist policymakers in the search and evaluation of the often-complex and controversial evidence. In light of this, the thesis also argues for engaging a variety of behavioral expertise and experts in various stages of the policymaking process. Failing to consider the controversial aspects of behavioral evidence and expertise could lead to policymaking that is inefficient, ineffective and which may even backfire.

The motivation for this thesis is not merely theoretical or philosophical, but also practical. Behavioral policies are the outcome of shifting from highly theoretical research to the interventions on decision-making in complex real-life situations. The thesis is theoretically directed in that it takes a closer look at the underlying empirical and normative assumptions of the evidence that is typically interpreted as justifying behavioral policies. It identifies and discusses several evidential challenges for behavioral policies, particularly related to agents’ own views on decision problems. The thesis is practically directed in that it aims to connect these evidential challenges to strategies for evidence-based policy to overcome or mitigate them. That is, it stresses the relevance of understanding the complex behavioral evidence in a nuanced way. It furthermore suggests ways to engage with the relevant expertise correctly. The thesis therefore consists of a mixture of both theoretical and applied approaches, connecting abstract theoretical chapters with either empirical or practical counterparts.

In order to sketch the theoretical context in which this thesis has been developed, the remainder of the introduction is structured as follows. First, I will provide a brief background in the history of behavioral policies. More specifically, the section zooms in on the historical context of research and theory development on decision-making and rationality. Second, I will discuss the concept of evidence-based policymaking. This includes a brief overview of some debates about the role of values and norms in science and evidence-based policymaking. Lastly, I will explain the structure of the thesis, and give a short preview of the chapters ahead.

1.2 A brief history of rationality & decision-making in economics

1.2.1 Expected Utility Theories

The foundations of “mainstream” or “standard” economics rest on what is called Expected Utility Theory (EUT). EUT holds that agents have complete sets of well-defined preferences over choice options. That is, agents always know which option they prefer over another, or whether they are indifferent between options. Agents compare options and prefer the option which brings the highest expected utility. Utility refers to the value of options multiplied by the (perceived) prospects of getting that value. So, in EUT, the value of for instance an item is not based on its price, but rather on the particular circumstances of the decision-maker and how much utility it brings her (i.e., marginal utility). Based on the utility of the options agents establish a preference ordering and rank their preferences over options from high to low (Bernoulli 1738/1954, Von Neumann and Morgenstern 1944). EUT was generally accepted as both a normative and a descriptive theory. That is, EUT was used as a theory to describe how agents make decisions, as well as a theory to prescribe how rational agents should make decisions. A rational agent, then, chooses the option which brings the maximum utility (e.g., Mongin 1998, Hausmann 2000, Briggs 2019).

At its core, EUT rests on mathematical theorems which Von Neumann and Morgenstern (1944) introduced for situations in which the probabilities over outcomes are given. Savage (1954) developed the subjective version of expected utility theory, which defines the probabilities in terms of preferences over acts. Savage's theory made it possible to measure subjective beliefs by probabilities. Crucial in either version of EUT are consistency requirements or 'coherence criteria' such as completeness, transitivity, and context independence. *Completeness* refers to agents' ability to rank all the choice options they face, thereby supposing that the agent has concrete preferences and also knows how to rank them from more to less preferred. That is, for any two choice options x and y , either the agent ranks x above y (prefers x to y), or the agent ranks y above x , or the agent is indifferent. *Transitivity* refers to the order of agents' rankings of the choice options. That is, if the agent prefers x to y and y to z , then the agent also prefers x to z . Famous arguments such as the money pump argument (e.g., Gustafsson 2022) show when preferences are intransitive, agents are likely to suffer harm. This is because if an agent's preferences are intransitive and she is willing to make exchanges, then she can be exploited. For instance, if an agent prefers x to y , y to z and z to x , and is willing to pay a certain amount of money EUR p to exchange y for x , z for y , and x for z , then, starting with z , she will pay EUR p for y , then EUR p again for x , then EUR p again for z and so on. *Context Independence* holds, roughly, that whether an agent prefers x to y does not depend on what other alternatives are available. For instance, if an agent prefers pears over apples, then her preference of pears over apples should not change when facing a choice between pears, apples, and bananas. However, it is crucial to realize that the descriptions of the choice options should contain everything that is relevant to the agent (e.g., Savage 1954, Hammond 1988). Only under those terms (amongst others) consistency is taken to be a criterion for rationality. After all, choosing ice cream over hot chocolate in summer and hot chocolate over ice cream in winter is not inconsistent if you take the relevant circumstances into account. Against this background, an additional criterion becomes important, namely description invariance. *Description invariance* implies that how the option is being described should not have an effect on the value or utility of the option. In other words, rational agents should not be influenced by the mere framing or presentation of an option. Rather, beliefs about options should be formed based on first order logic and Bayesian probability (e.g., Maher 1993, Tversky and Kahneman 1986).

The wide acceptance of EUT in economic theory is due to a variety of aspects. First, the theory captures a formal rather than a substantive notion of rationality. Agents choose according to what they perceive the utility of certain options and outcomes are. EUT prescribes how rational choices and preferences should be consistent and coherent but makes no claims about what valuable goals for agents are. It thus allows for a wide variety of *preferences*. Second, EUT allowed the possibility to measure behavior by linking preferences through choice observations. One interpretation is that the statement that an agent strictly prefers bundle x to bundle y means nothing more than that the agent systematically chooses x over y . Another interpretation is that the statement *in addition* implies that the agent judges himself to be better off with x than with y (Mandler 2001). Third, and relatedly, EUT made it possible to measure welfare in a numerical way through the concept of utility. Expected utility theory thereby proved to be crucial for the emergence of behavioral economics which entailed experiments showing systematic deviations from (subjective) expected utility theory, as we will see below (see also Nagatsu 2015, Moscati 2016). Despite the theory's advantages, several

economic scholars criticized the normative acceptability and empirical plausibility of (subjective) expected utility theory (e.g., Markowitz 1952, Allais 1953, Ellsberg 1961, Sen 1977). These criticisms were mainly based on thought experiments, as experimental economics was not yet common scientific practice (Nagatsu 2015). For further background on EUT and its reception, see for instance Bradley (2017), Mongin and Baccelli (2021), Nagatsu (2015), Moscati (2016) and Malecka (2020).

1.2.2 Bounded Rationality

Around the 1970's, behavioral decision researchers became dissatisfied with what is often referred to as the 'standard picture of rationality' in decision theory (Stein 1996, Polonioli 2016). Before that time, refutations of EUT such as by the Allais and Ellsberg paradoxes were mostly treated as anomalies. However, experimental testing became more common in economics and experiments generated overwhelming evidence that agents do not behave according to the standards of EUT. More precisely, the experiments show that EUT was violated systematically. Part of these experiments were for instance empirical experiments based on the Allais paradox and other experiments showing 'preference reversals', such as those by Lindman (1971), and by Slovic and Lichtenstein (1968) (see Seidl 2002). The idea behind these experiments was that agents' preferences can be elicited by various methods. For instance, if agents can be asked to state their preferences over various lotteries, then their preferences can be derived from observing which lotteries they choose amongst multiple ones, or they can be asked to state prices they would be willing to pay to play. Rational agents, or so it was assumed, should show the same preferences in each of these cases. However, different methods showed different preferences which made various scholars such as these question the validity of EUT. While these scholars directed their attention mainly to specific criteria for rationality of EUT such as independence, Simon (1955, 1996) offered a more fundamental critique by introducing *bounded rationality* (Nagatsu 2015). According to Simon, the standard rationality assumptions should be replaced with criteria that do justice to the limited cognitive capabilities of human beings, who have limited access to information, in the daily environments in which decisions have to be made (Simon 1955a:99). In that sense, Simon proposed a "cognitive revolution" which entailed a shift (back) from the focus on choice behavior towards the modelling of cognitive internal processes and reasoning (Heukelom 2010, Nagatsu 2015).

Regarding the descriptive ambitions of EUT, Simon argued that EUT lacked predictive power because it focuses on utility maximization rather than what he dubbed 'satisficing'. Agents have limited resources to make decisions, in terms of cognitive capabilities, information and time. Rather, agents make a tradeoff between the effort put in the decision-making and the accuracy of the decision. For instance, Simon argues that agents do not compare all possible alternative choice options, then calculate the expected utility and choose the most optimal option. Instead, agents choose the first option that satisfies their 'aspiration level'. Crucial in this analysis is the structure of the environment in which the agent makes decisions. Simon is interested in human beings making decisions 'in the wild'. In Simon's conceptualization, ecological structure contains both biological features of humans and the task environment (Simon 1955). The project of bounded rationality then, focused on researching deviations from the standard model of rationality. In addition, it also sought to develop new methods to model decision-making in economics which came closer to the describing actual choice behavior. Thereby, it developed new, more realistic theories about agents' cognitive processes and their perspectives on the decision problem.

Regarding the normative aspects of EUT, Simon argued that, in many environments, maximization is not the most rational process. Instead, “satisficing” is (Simon 1955). In this context, Simon’s bounded rationality project also entails re-evaluating and reforming the concept of rationality (e.g., Simon, 1955, 1976, Grüne-Yanoff 2007, Heukelom 2010, Nagatsu 2015).

Simon’s work on bounded rationality has inspired two different, competing, research programs. One is the Heuristics and Biases program within the behavioral economics tradition. The other is the Simple Heuristics program within the ecological rationality tradition. Both research programs claim to draw on the concept of bounded rationality as proposed by Simon. However, they do so based on different methodological principles and experiments. They also draw on competing concepts of rationality (Samuels, Stich and Bishop 2012). The behavioral policies subject to this thesis are often advocated as based on the experimental research and normative theories developed within these research programs. They are competing behavioral policies developed based on the Heuristics and Biases program and the Simple Heuristics program respectively. The following two sections briefly discuss the Heuristics and Biases program and Simple Heuristics program respectively. After that we discuss the behavioral policies from these programs.

1.2.3 Behavioral economics & the Heuristics and Biases Program

1.2.3.1 Kahneman and Tversky

While experimental economics had already taken off, Kahneman and Tversky (e.g., 1979, 1986, 1989, 1992) brought experimental research on human decision-making to a new level. Quintessential about the work of Kahneman and Tversky, and one of the main factors for their success, is a distinction between the requirements of rationality as descriptive features and normative conditions (Heukelom 2007, 2014, Nagatsu 2015). Roughly, they saw bounded rationality as a concept explaining agents’ decision-making restraints, but not as a different normative standard for rationality. Kahneman and Tversky argued that economics had used the requirements to serve both a descriptive and normative goal. While they argued that the requirements for rationality were still apt from a normative perspective, they demonstrated empirically that they were not descriptively accurate (e.g., Kahneman and Tversky 1979, Kahneman and Thaler 1991).

For instance, various heuristics and biases experiments showed that agents are loss-averse to a high extent. These findings were often accompanied by agents’ tendency to be more risk averse when presented with gains and show more risk-taking behavior when presented with losses. This implies that agents place a higher value on losses than on gains. However, also the mere description of a choice in gains or losses could trigger loss aversion. A now classic example of such a framing case in which agents violate transitivity in a systematic manner is the Asian disease experiment (Tversky and Kahneman 1981).

In the Asian disease experiment subjects were presented with two choices:

Choice A:

Program 1: 200 people are saved

Program 2: 600 people are saved with $p=1/3$, no one is saved with $p = 2/3$

Choice B:

Program 3: 400 people die

Program 4: No one dies with $p = 1/3$, 600 people die with $p = 2/3$

Studies show that approximately 75% of the subjects prefer option 1 to 2 in choice A, whereas in choice option B 4 is preferred over choice option 3 by 75% as well. However, programs 2 and 4 are different descriptions of the same choice. Saving 600 people with $p1/3$ is logically the same option as not killing 600 people with $p1/3$. And saving 200/600 people is the same option as killing 400/600. These empirical findings are generally analyzed as showing an apparent inconsistency, more particular a violation of transitivity, brought about by mere framing of the wording of the options.

In addition to demonstrating systematic deviations from EUT, Kahneman and Tversky also combined Tversky's theoretical framework with Kahneman's work in psychology on cognitive mistakes. Consequently, Kahneman and Tversky did not only provide a framework that showed how agents diverged from the norm. They could also offer explanations of why they did so (Heukelom 2007). Tversky and Kahneman (1974) argue that agents suffer from several biases and furthermore use unreliable heuristics (i.e., mental shortcuts) to tackle a decision problem. In the Asian disease case, for instance, Kahneman and Tversky (1982) argue that agents suffered from the bias of loss aversion. This results in agents placing a higher value on avoiding a loss than acquiring a gain. Furthermore, agents use faulty heuristics. This led them to fail to understand that the description of the decision problem is irrelevant, or so Kahneman and Tversky argue. Examples of heuristics that are often referred to in the context of framing effects are the affect heuristic (i.e., relying on an emotional state to make decisions) and the availability heuristic (using information that comes to mind quickly.) Kahneman's and Tversky's line of work to systematically predict and explain deviations from standard rationality later became known as the 'Heuristics and Biases research program' (Sent 2004). Often, researchers within the Heuristics and Biases program explain framing results with reference to a dual process system or dual system approach of decision-making. This dual system approach models human cognition as an automatic system invoking biases and faulty heuristics (system 1) and a more reflective cognitive system which invokes more rational reasoning (system 2) (e.g., Stanovich 1999, Kahneman and Frederick 2002, Kahneman 2011, Thaler and Sunstein 2008, Vlaev et al. 2016).

1.2.3.2 Prospect Theory

Kahneman and Tversky used the deviations from standard criteria of rationality as demonstrated in experiments to develop prospect theory (Kahneman and Tversky 1979, 1992, 2013). According to prospect theory, agents place less weight on outcomes that are merely probable than on outcomes that are certain, which implies so-called 'risk aversion'. Moreover, as discussed, agents are loss averse, which means that agents prefer avoiding losses over acquiring gains. Prospect theory takes these findings into account by providing a model under which agents behave as if they assign a utility function to the outcomes. Subsequently, they choose the outcome with the highest utility. The model contains two main steps: *editing* and *evaluation*. In the editing stage, agents decide between which outcomes they are indifferent, use that as a reference point and then assign to the outcomes the predicates 'loss' or 'gain'. Subsequently, in the evaluation phase, agents behave *as if*

they assign a utility function to the outcomes and then choose the outcome with the highest utility. Prospect theory thereby not only explained framing effects but also theoretical problems such as assigned values to gains and losses rather than to final outcomes. Even though theories of bounded rationality such as prospect theory are primarily descriptive, they can be applied normatively as well. For instance, it could be argued that it may be rational to reduce the cognitive burden of decision-making by employing certain heuristics.

The experimental findings which showed that agents deviate from the standard picture of rationality in a systematic manner were widely accepted in economics. However, the conclusions drawn from it by the Heuristics and Biases program are not. For instance, the deviations spurred many questions about the normative value of EUT generally (e.g., Gigerenzer 1991, Hertwig and Gigerenzer 1999, Gigerenzer and Brighton 2009, see also Vranas 2000). If agents violate these criteria so systematically, then why hold on to standards that seem impossible to achieve? (see also Basel and Bruhl 2012). Furthermore, scholars argued that experiments were often not set up adequately. They triggered certain behaviors that were not representative for behavior outside of the lab. Also, agents may not have perceived or understood the choice options in the way that the experimenters analyzed (e.g., Gigerenzer 2015, Cosmides and Tooby 1996, Binmore 2008, 2009, Plott 1996). Relatedly, more general critique was given about the descriptions of choice options in the (normative) decision analysis in mainstream economics. When analyzing behavior on for instance consistency, experimenters did not describe the choice options with all aspects that were relevant to agents, or so it was argued (e.g., Sen 1977, Broome 1991).

1.2.4 Ecological Rationality & the Simple Heuristics Program

A foundational and programmatic critique on the Heuristics and Biases program came from other bounded rationality scholars working in evolutionary psychology. They argue that the Heuristics and Biases program did not interpret Simon's concept of bounded rationality correctly. With Gerd Gigerenzer as its main advocate, a research program based on the concept of adaptive or ecological rationality emerged (Gigerenzer 2000, Todd and Gigerenzer 2012). Ecological rationality doubles down on Simon's idea that the match between agents' cognitive structures and the environment in which decisions need to be made is crucial to evaluate rationality. Building on his 1955 paper on bounded rationality, Simon (1956) stressed that, if the right environmental cues are attended to, cognitive decision-making decision processes can be quite minimal. In that light, proponents of ecological rationality argue that the behavioral experiments did not only give rise to questions concerning the descriptive value of the standard picture of rationality. It also prompted questions about the normative value of the (rationality) criteria for consistency. Ecological Rationality is based on the empirical findings that simple strategies can sometimes be as good as or even better than optimizing strategies that need more information and computation (e.g., Gigerenzer and Brighton 2009, Gigerenzer 2021). Rather than focus on so-called *coherence or consistency criteria* as the standard picture of rationality does, it takes so-called *correspondence criteria* as a normative benchmark for rationality. Correspondence criteria include the extent to which one can make accurate predictions, successful exchanges with others, and the organism's survival and reproductive success (Gigerenzer and Todd 1999, Marewski, Gassmaier and Gigerenzer 2010, Todd and Gigerenzer 2000). According to advocates of ecological rationality, these criteria are minimal and not intended as universal criteria for rationality. Ecological rationality holds that there may be specific environments in which these criteria lead to success, but also many environments in which they

are either unimportant or in which it is even rational to violate them (e.g., Berg and Gigerenzer 2006, Arkes, Gigerenzer and Hertwig 2016). Relatedly, scholars from the ecological rationality program differ fundamentally from the Heuristics and Biases program in their normative understanding of heuristics. According to ecological rationality scholars, heuristics are often not irrational or suboptimal decision-making processes. Heuristics may be outdated and not fit anymore for the situation in which agents apply them. However, generally they are decision-making processes that have been ecologically adapted to help make decisions under constraints. If heuristics are well suited to their environments, then using heuristic is an advantage in decision-making rather than a flawed or suboptimal reasoning process. They are ‘simple heuristics’ (i.e., “fast and frugal”) and could help agents make efficient and accurate (enough) decisions (Gigerenzer and Todd 1999, Gigerenzer 2004, Hoffrage and Reimer 2004, Hafenbrädl et al. 2016). Here it should be noted that, recently, advocates of the Heuristics and Biases program also acknowledge that heuristics are not fundamentally irrational (e.g., Sunstein 2017). However, as we will discuss in chapters 2, 4 and 5, in the context of behavioral policies they still are conceived as such.

Besides the normative critique of the standard criteria of rationality, the ecological rationality program also distinguishes itself in its empirical endeavors. First, there is a sharp methodological difference. The ecological rationality program aims to specify conditions under which heuristics work better than competing decision-making processes (e.g., Berg and Gigerenzer 2010). This is sometimes also referred to as the Simple Heuristics program (Grüne-Yanoff and Hertwig 2016). Heuristics, in their environments, are treated as the main object of study themselves, rather than as explanations of behavioral effects such as in the Heuristics and Biases program (e.g., Todd and Brighton 2016, Goldstein and Gigerenzer 2002). In that sense, Simple Heuristics scholars argue that heuristics can also be understood as computational methods. They criticize the Heuristics and Biases program about for keeping heuristics too vague, which renders them unable to explain how and when they operate (e.g., Todd and Brighton 2016, Berg and Gigerenzer 2010). The different kind of research methodology also leads to different interpretations and understandings of the behavior. According to Simple Heuristics scholars, there are different explanations for the observed behavior changes. This will be discussed further in chapters 2 and 4. Second, and relatedly, the Simple Heuristics program does not endorse the dual system approach. For instance, according to Gigerenzer and other scholars, the use of heuristics is not necessarily automatic. Heuristics can be applied both deliberately and non-deliberately. Relatedly, they also argue that human cognition cannot be categorized as consisting of two systems (Kruglanski and Gigerenzer 2011, see Grayot 2020 for an overview).

1.2.5 From theoretical programs to behavioral policies

So far, I presented a cursory history of decision-making and rationality (for more comprehensive overviews, see Van Heukelom (2014), Nagatsu (2015), and Sent (2004); Angner and Loewenstein (2012) offer an informative review). For now, it is important to state that both the theoretical and experimental ‘legs’ of the heuristics and biases program and the Simple Heuristics program gave rise to the development of practical interventions. That is, based on the Heuristics and Biases program and the Simple Heuristics program two different, competing, types of behavioral policies have been developed.

We have seen that both research programs depart from different empirical and normative assumptions, have different methodologies and furthermore experimental results. In other words, there is much scientific disagreement about the evidence for behavior change, and its empirical and normative interpretation. This raises the question central to this thesis to what extent behavioral policies are evidence-based. In the following section, I will sketch the aims and ambitions of evidence-based policy and behavioral policies specifically. Furthermore, I will provide a brief overview of debates concerning expert disagreement and the role of norms and evidence more generally.

1.3 Evidence-Based Behavioral Policies

1.3.1 Evidence-based behavioral policies

Evidence-based policy aims to develop policies that are backed up with the ‘best evidence’ (e.g., Davies, Nutley and Smith 2000, Sanderson 2002, Head 2008). Although there is not one widespread definition of evidence-based policy, it generally holds that policies are based on extensive or high-quality experimental research such as randomized control trials, and meta-analyses (Boaz, Ashby and Young 2002, Young, Ashby and Boaz 2002). This is in contrast with developing public policy based on ideology or “common sense” ideas about which policies work and why they do so. The aim of evidence-based policy is to develop policies that are more efficient, effective and robust in the long term (e.g., Cartwright 2012). A popular phrase used in evidence-based policy practice to describe its main ambition is to implement “what works”. For instance, in the UK many so-called What Works Centres or evidence-based policy teams have been installed, each drawing on specific evidence-based policy methods (e.g., Behavioural Insights Team 2012, 2016). Often, evidence-based policy works with evidence hierarchies in which quantitative evidence is placed over qualitative evidence (e.g., Cartwright, Goldfinch and Howick 2009, Clarke et al. 2014, Kosrowhi 2019). In particular, randomized control trials (RCT’s) are considered superior over other types of evidence. To put it even more strongly, RCTs are often promoted as “the gold standard” for evidence-based policymaking (e.g., Cartwright 2009).

The “evidence-based policy movement” has been catalyzed by the popularization of the behavioral insights found in behavioral economics and psychology (e.g., Marchionni and Reijula 2019, Rostain 2000). Books such as *Thinking fast and slow* (Kahneman 2011), *Predictably irrational* (Ariely 2008), and *Nudge* (Thaler and Sunstein 2009) have brought scientific insights from academia to the broader public, including policy makers. Although there is not one widespread accepted definition of behavioral policies and they come in many shapes and forms, they can be roughly defined as policies aiming to invoke behavior change based on the psychological insights about human decision-making. The idea behind behavioral policies is that small tweaks in the presentation of choices thus can be used to improve individual and social outcomes and do so at relatively low cost.

1.3.2 Nudge, Boost and combinations

There are many types of behavioral policymaking (e.g., Michie et al. 2015, Grüne-Yanoff, Marchionni and Feufel 2018), although often the term refers to one particular type: Nudge. Nudge originated from the Heuristics and Biases program. Drawing on Rebonato (2012, p. 32) and Grüne-Yanoff and Hertwig (2016), Nudge can be defined as consisting of the following characteristics:

1. *A Nudge is intended by the policy maker (choice architect) to steer the chooser's behavior away from the behavior implied by the cognitive shortcoming and toward her ultimate goal or preference (e.g., healthier food choices).*

2. *A Nudge seeks to realize this influence by exploiting empirically documented cognitive shortcomings in human deliberation and choice, without changing the financial incentives (disincentives).*

3. *A Nudge does not affect those features over which people have explicit preferences (e.g., money, convenience, taste, status, etc.), but rather those features that people would typically claim not to care about (e.g., position in a list, default, framing).*

4. *The behavior change brought about by the Nudge should be easily reversible, allowing the chooser to act otherwise.*

Famous examples of Nudges are opt-out systems for organ donation, saving programs that ask employees to save from a future salary increase rather than their current salary, and 'framing' by means of the wording of choice options in positive or negative terms (e.g., Wilson, Kaplan and Schneidermann 1986, Lynge 1998, Thaler and Benartzi 2004, Johnson and Goldstein 2013, Nisa et al. 2019). Nudges are often modelled after RCT's conducted within the heuristics-and-biases research program and the results of these experiments are explained by high-level meta-analyses on the mechanisms (potentially) bringing about these results. Therefore, Nudge is often advocated as evidence-based. Furthermore, Nudges are often implemented and advocated through "What Works Centres" and behavioral insight teams (e.g., Behavioural Insights Team 2012, 2016, Bristow, Carter and Martin 2015, Ball and Head 2020).

Despite its popularity, Nudge also faces severe criticism, much of which concerns the ethics of implementing them. Criticisms for instance regard bypassing people's autonomy, lack of transparency and the political ideology through which Nudges are often referred to, namely "libertarian paternalism" (Thaler and Sunstein 2003). Libertarian Paternalism is a version of soft paternalism. It holds that since agents are always influenced by the presentation of options in one way or another, choice options should be presented in a way that help agents be "better off" (Thaler and Sunstein 2003, 2009). Critics often argue that libertarian paternalism is mostly paternalistic, because it claims to know what outcomes make agents better off and circumvent rational decision-making processes by invoking automatic ones (Bovens 2009, Hausmann and Welch 2010, Grüne-Yanoff 2012, Rebonato 2014).

Partly as a response to criticisms on Nudge and the Heuristics and Biases program more generally, an alternative policy named Boost has been developed within the Simple Heuristics program. According to Grüne-Yanoff and Hertwig (2016, 2017), Boosts have the goal of expanding people's competences to reach their own objectives without presupposing or having assumptions about what these goals are. They distinguish three classes of Boosts:

- (1) change the environment in which decisions are made,
- (2) extend the repertoire of decision-making strategies, skills, and knowledge,

or (3) do both. (Grüne-Yanoff and Hertwig 2016, p. 156)

Famous examples of Boosts are fast and frugal decision trees that replace lengthy diagnostic questionnaires for healthcare workers (e.g., Jenny et al. 2013), statistical training (e.g., Gigerenzer and Hoffrage 1995, Sedlmeier and Gigerenzer 2001), and changing the representation of statistical information from relative to natural frequencies (Gigerenzer et al. 2007, Kurz-Milck, Gigerenzer and Martignon 2008). What all these Boosts have in common, is that they strive towards creating an adequate fit between agents' decision-making competences and a specific choice environment. That is, Boosts aim to present choices, and information more generally, in such a way that they are easy to understand and act upon, without steering into the direction of one particular choice option. In that sense, Boosts focus on beliefs rather than on choices themselves, even though the aim is that better beliefs eventually lead to better choices. Boost advocates therefore also contradict the argument from Nudge advocates that because 'choice architecture' is unavoidable, agents are always necessarily steered into one or another direction. Their research shows that, for instance, behavior change effects disappear once choice options are described in both positive and negative wording or both gains and losses. This in contrast to Heuristics and Biases research in which experiments only describe the choice options in one way (e.g., Gigerenzer 2015).

Besides disagreements on the aims, justification and evidence for Nudge and Boost as competing policies, they are not necessarily mutually exclusive. Combinations are possible (e.g., Hertwig and Grüne-Yanoff 2017). For instance, Nudges and Boosts can be combined in saving application forms in which both an interactive calculator to boost understanding of saving amounts and a social commitment pledge form (Nudge) are integrated amongst other features (Timmons, Robertson and Lunn 2022). Furthermore, there are alternatives to Nudges and Boosts such as 'active choice design' in which individuals have to make a choice (e.g., Carroll et al., 2009, Keller et al. 2009). An example of this is only giving out driver's licenses if agents have actively chosen between registering as organ donors or not (e.g., Stevens et al. 2019). Hence, behavioral policies include a wide range of policies drawing on behavioral economic and psychological insights. In several chapters of the thesis Nudge and Boost and their theoretical underpinnings will be discussed in more detail (in particular, chapters 2 and 4).

1.3.3 Critiques on behavioral policies as evidence-based policies

In the previous sections, we have discussed the competing heuristics and biases program and the simple heuristics program, and Nudge and Boost as alternative behavioral policies. Research from both programs can be understood as critiques on each other's goals, methodologies, and evidence-base. However, there also is more general critique from the philosophy of science about behavioral policies. For instance, one of the most often voiced critique by philosophers of science is that much behavioral evidence stems from RCTs. According to them, RCTs only show statistical evidence rather than causal evidence (e.g., Russo and Williamson 2007). Therefore, they claim that the evidence for evidence-based policymaking should be generated by a combination of different types of methods, such as a combination of RCTs and more qualitative research (Cartwright and Hardie 2012, Dede 2019). This would also help in understanding why the behavior change only occurs under some particular conditions (e.g., Marchionni and Reijula 2019, see also chapters 2,4 and 5). Furthermore, there is a debate about the relation of Nudge and Boost with the research programs they originated from. For instance, Grüne-Yanoff and Hertwig argue that the connection between

Nudge and the Heuristics and Biases program is often weak and that, to a lesser extent, the same holds for Boost and the Simple Heuristics program. Furthermore, there is critique about Nudges steering into directions that are aimed to ‘help people make decisions as judged by themselves would they have been fully rational’ (Thaler and Sunstein 2016). This line of critique is based on the findings of the heuristics and biases program that shows that it is often unclear whether people’s choices show preferences, and that there is not a reliable preference elicitation method yet (Bovens 2009, Infante, Lecouteux and Sugden 2016, see Sugden 2022 for a nuanced analysis of recent methods in this light). Chapters 2 and 4 of this thesis in particular focus on the role of theory for the evidence-base of behavioral policies.

In addition, there are also more general concerns and criticisms about the aims and feasibility of evidence-based policy. For instance, in many policy areas the research base is not strong enough to be adequately informing policy practice. Also, policymakers often have to consider many other factors besides research evidence, such as speed, stakeholder management, risk management, and external perception. Furthermore, there often is not a good fit between the research topics, methods, and dissemination from the academic world and the needs of the practical policy world. Relatedly, there is an increasing focus on engaging the expertise of non-academic professionals. However, how to integrate this professional expert knowledge with academic expertise is still unclear). Last but not least, academic knowledge seems to have less standing and is trusted less in situations where the issues are subject to rapid change (e.g., Head 2010, Sanderson 2002, Newman 2017). Hence, evidence-based policymaking is not by any means “straightforward” or uncontroversial. The evidence-base for behavioral policies is no exception. In fact, behavioral policies may be even more controversial because of the deep scientific disagreements about their empirical and normative assumptions, or so I will discuss in chapters 2 and 5. The following section briefly discusses debates in the philosophy of science about the role of values and norms in scientific evidence and evidence-based policy, including issues concerning expert disagreement.

1.3.4 Values in science

Behavioral policies and the normative interpretations underlying them prompt a crucial question from an important general debate in evidence-based policy. That is the question to what extent values can and should play a role in science and evidence-based policy specifically (see also Dede 2021). While this thesis does not directly contribute to the values in science debate, the literature does play an important role in the background to which this thesis has been developed. That is, this thesis discusses the often complex and conflicting scientific evidence for behavioral policies. Since the experiments are done and interpreted in the light of the normative rationality framework, merely additional experiments (even of a different type) will not be the sole solution. Rather, rationality assumptions underlying the interpretation of experimental results must be made explicit and various values relevant to policymaking must be considered.

An important line of debate in the values in science literature stems from Rudner (1953) who argues that values are necessarily involved in scientific hypothesis acceptance. Determining the threshold for ‘sufficient’ evidence for confirming or rejecting the hypothesis involves ethical value judgements about the importance of making a mistake in accepting or rejecting a hypothesis. For example, some individuals may die of the side effects of a drug mistakenly judged to be safe. Others may die of a condition because they did not have access to a treatment that was mistakenly judged

to be unsafe. Rudner's account has been criticized for failing to distinguish between more fundamental research, where these direct ethical concerns may not play a role, and applied research with direct practical implications such as he describes. Despite the criticism, Rudner's argument has been accepted widely in a modified form. For instance, Douglas (2000) argues that since even strongly confirming evidence cannot fully prove a universal scientific law, we must live with a residual 'inductive risk' in inferring that law. Contextual values influence scientific methods by determining the acceptable amount of inductive risk. She furthermore argues that although the acceptance of scientific theories is an especially prominent and explicit one in which values enter the scientific process, it is only one of several places. Many decisions in the process of scientific inquiry may conceal implicit value judgments: the design of an experiment, the methodology for conducting it, the characterization of the data, the choice of methodology for processing and analyzing data, the interpretational process findings, and so forth. None of these decisions could be made without consideration of the possible consequences that could occur. Furthermore, she argues that expert disagreement ultimately boils down to disagreement about the sufficiency of evidence. Of course, differences in experts' training and experience will generate different perspectives on many issues, including potential pitfalls in evidence and the importance of different cognitive values such as clarity and conciseness. However, in communicating the evidence experts generally are and should be aware of the disagreements and take them into account in their evaluation of the sufficiency of evidence. Steel (2010) defends a similar point about the inevitable role of values in science by analyzing the role of scientists as policy advisors more explicitly. She builds on Rudner's account (1953) and argues that scientists often for pragmatic reasons must communicate their findings on a cruder plausibility scale than appropriate for representing their beliefs. In science communication for policy a form of translation is always necessary. Like Douglas, he also discusses how this extends beyond the information about the truth of the hypothesis, such as advice on the refinement of the policy problem and advice on whether more evidence needs to be collected (2010, 2016). Furthermore, various scholars argue that researchers have a moral responsibility to consider the (policy) consequences of their work even when they merely make empirical claims and refrain from policy advice. After all, research may be picked up by policy makers through different, indirect, channels (e.g., Douglas 2000, 2009, Gundersen 2018).

However, the idea that scientists *necessarily* make value judgements has been disputed by many scholars in the philosophy of science. Betz (2013), for instance argues that scientists can largely avoid making contextual value judgments if they carefully express the uncertainty involved with their evidential judgments, e.g., by using a scale ranging from purely qualitative evidence (such as expert judgment) to precise probabilistic assessments. Furthermore, scholars argue that *even if* it is impossible to separate facts from values in science, then scientists should still strive for objectivity in science to provide an "antidote to individual and institutional interests" (Lacey 1999, 2002, McMullin 1982, Mitchell 2004). A related general worry is that if (non-epistemic) values indeed play an important role in science communication and policy advice, then scientific experts take partly the role of those who are democratically elected (e.g., Steel 2010, Kitcher 2011). Some scholars develop 'solutions' to mitigate this. One example is Douglas's 'detached objectivity', according to which it is prohibited for values to replace or dismiss scientific evidence. Douglas distinguishes indirect role of values regarding inductive risk which she opposes to direct roles of values when they amount to the refutation or acceptance of evidence. According to her, the key norms for experts are to keep values out of the direct role and to make values in the indirect role explicit

(Douglas 2004). Furthermore, Kitcher's well-ordered science account prescribes that only those scientific inquiries (from agenda setting to its applications) are pursued if a well-informed, representative, and cooperative group of deliberators would have chosen them (Kitcher 2003).

Yet another, very different, line of argument is that values are beneficial for good scientific practice. A value-laden science can also meet important criteria for being epistemically reliable and objective (e.g., Anderson 2004; Kourany 2010, Longino 1990, 1996). Longino has been especially influential in assigning a crucial function to social systems of criticism in securing the epistemic success of science. She conceives scientific knowledge as a product of social interactions and a diversity of values. Specifically, she develops an epistemological account which regards a method of inquiry as "objective to the degree that it permits *transformative criticism*" (Longino 1990, p. 76). Also other scholars argue that disagreements have important epistemic values: heterogenous communities of scientists who use different methods and have divergent interpretations of data generally function better and collect more and higher quality evidence (e.g., De Cruz and de Smedt 2013).

In the context of behavioral policymaking, the discussion about values in science prompts the question: does acknowledging that value-laden research is relied on in behavioral policymaking challenge the idea of behavioral policies being evidence-based? Arguably, following the line of Douglas, Betz and others who claim that scientists must try to avoid making value judgements or mitigate them, the answer would be negative. However, this would be the case only provided that values which are unavoidable are being made explicit. Regarding rationality evaluations, this could entail making the rationality standards for behavioral policies and the underlying evidence explicit or disregarding them altogether. Moreover, scientific disagreement on empirical and normative assumptions of behavioral research would need to be considered. This would be especially important in the context of judging whether the evidence is sufficient to justify behavioral policies. Furthermore, in line with Kitcher, a crucial condition for behavioral policies to be pursued is if a representative group of deliberators agree. Also for scholars who see values in science as something that needs to be harnessed rather mitigated such as Longino (1990), a social system needs to be in place. Criticism and disagreement should take center stage in develop evidence-based policies. Making the scientific disagreements explicit could also assist in building a stronger body of evidence. It goes beyond the scope of this thesis to analyze and apply the different arguments to behavioral policies in detail. However, it is important to note that whatever stance one takes on the role of values in science, these values need to be made explicit to see whether they are (un)avoidable, how they can be mitigated, or harnessed. This thesis aims to contribute to making empirical and normative disagreements about behavioral evidence explicit. More specifically, it focuses on the rationality evaluations of people's behavior and the implications for the justification of behavioral policies as evidence-based.

1.4 Preview of the chapters

The remainder of this thesis is structured as follows. The thesis consists of two parts, each containing one theoretically oriented chapter and one more practically oriented chapter. The chapters have been conceived of as self-standing research articles.

Part 1 focusses on the interpretation of *behavior change* demonstrated in a wide variety of behavioral experiments. The interpretations for behavior change in the experiments often draw on theoretically contested assumptions, or so I will argue. Specifically, these assumptions concern agents' views on the choice options, and an evaluation of whether or not these views are brought about by rationally justifiable processes. These assumptions are rooted in abstract decision theoretical research. Furthermore, qualitative research shows that people's views on choices such as organ donation registration are more complex and nuanced than assumed. This also has implications for behavioral research. For instance, a combination of quantitative (e.g., RCT) research and more qualitative research can get more insight in why people may engage in behavior change as a result of, for instance, defaults. Furthermore, it has implications for the behavioral policies themselves. It may show that although behavioral policies may seem effective, it is not in the long term or they do not increase welfare benefits.

Chapter 2 discusses how exactly behavioral policies such as Nudges are based on randomized controlled trials (RCTs) and lab experiments which show that the way a choice is presented influences agents' behavior. I argue that behavioral economics and Nudge rely on a specific interpretation of the experimental results as irrational inconsistencies. I show that this interpretation itself implicitly presupposes evidence for agents' cognitive processes and beliefs about the choice options. Consequently, the evidence for behavioral policies should be expanded with research that shows how agents view the decision problem in front of them. Without this, the evidence for behavioral policies neither establishes internal nor external validity. That is, the evidence will neither show that agents will show behavior change in new policy environments, nor that they did in the experimental set-ups in the first place. Moreover, the chapter shows that the normative role of agents' cognitive processes and beliefs in the analysis of the decision problem is theoretically contested. This poses what I call the 'individuation challenge' for behavioral policies: to establish a sufficient evidence-base, more empirical research needs to be done on people's own perspectives on the choice options, and a normative foundation should be developed for a theory of options. Chapter 2 is currently under revision as a research article for the journal *International Studies in the Philosophy of Science*.

Chapter 3 explores agents' views on a much-discussed behavioral policy case: organ donation registration. In many countries, such as the US, Germany, France, and the Netherlands, governments are dealing with a great shortage of organ donors. Even though people generally show positive attitudes towards organ donation, they often do not actually register as organ donors themselves. The objective of the study discussed in the chapter was to explore prevailing viewpoints among the Dutch population on deceased organ donation and the relation between aspects of the viewpoints potentially influencing the decision to register as an organ donor. Although substantive research about attitudes on organ donation has been conducted, this is the first study investigating people's viewpoints focusing on the *relation* between beliefs, tastes, preferences, motives, goals, and other constituents underlying people's viewpoints on organ donation, such as the role of the media and public policies. The study uses a mixed method approach combining both qualitative and quantitative methods called Q-methodology. The Q-methodology study revealed four existing viewpoints: "not donating your organs is a waste", "it does not go with my religion", "my family should decide later"; and "it's a good deed, but I'm doubtful". These viewpoints convey information on potential reasons for the gap between people's

favorable attitudes towards organ donation and the low number of actual registrations. It also implies that while defaults can help increase organ donor registrations for some target groups, these target groups either are in doubt or have their family members decide later on. Hence, an increased number of registrations ultimately may not lead to an increased number of actual donations. Chapter 3 is based on a research article that has been published in *Plos One* (Truijens and van Exel 2019).

Part 2 of the thesis moves to the question what the implications of the theoretical (dis)connections of behavioral policies are for *evidence-based* policymaking and policymakers. It discusses several evidential challenges and explores how and if policymakers can deal with them. I argue that in order to develop effective behavioral policies, policy makers should be aware of the evidential challenges and engage with the evidence cautiously. In order to support this, I explore and offer suggestions for policymakers to do so.

Chapter 4 discusses the question to what extent “theory policy coherence” is important for evidence-based behavioral policymaking. Based on Grüne-Yanoff and Hertwig (2016), I argue that evidence for behavioral policies often conflicts with evidence from rivaling theoretical programs. Furthermore, even the connection between the behavioral policies Nudge and Boost and the research programs they originated from is often weak. I argue that this is crucial in two ways: first, it is crucial for the evidence-based policy maker to realize this, as assuming theory-policy coherence while it does not exist or is weaker than assumed can lead to an incorrect assessment of evidence and ultimately to adoption of policies on false grounds. Second, the concept of theory-policy coherence may assist the policy maker in the search and evaluation of (mechanistic) evidence. However, it is crucial to realize that it only does so provided that the policy maker takes into account the complexities of the relations between the theories and policies in question. Chapter 4 is based on a research article that has been published in the journal *Rationality & Society* (Truijens 2022).

Chapter 5 brings the evidential challenges for behavioral policies identified in this thesis and in the philosophy of science generally to the policymaker. Furthermore, it suggests alternative applications of the toolboxes currently used by behavioral policy makers. More specifically, in the chapter, I argue that the evidence for behavioral policies is often not as strong as it seems. The evidence shows that people change their behavior after seemingly irrelevant changes in their environment, but not why they do so. And the evidence may show that people behave irrationally, but only based on assumptions about people’s views of the choice options and according to very specific and highly debated theories about rationality. I argue that policy makers who are not aware of these nuances may develop policies based on one-sided expertise and evidence. This can lead to policies that are inefficient and ineffective, and which may even backfire. This chapter therefore brings policymakers up to speed with evidential challenges and argues for a nuanced view of behavioral evidence. In addition, it presents a framework for adjustments to the toolbox of policy makers in order to navigate the complex, and often controversial, behavioral evidence. In short, the chapter recommends policy makers to include scholars with theoretical and empirical expertise from different research traditions in expert panels, develop so-called evidence gap maps or systematic literature reviews including the policy's assumptions and the evidence for and against

them, and evaluate (non-behavioral) policy alternatives. An applied research article based on Chapter 5 has been submitted at the journal *Behavioral Science and Policy*.

Chapter 6 concludes and summarizes by discussing the evidential challenges and implications for evidence-based behavioral policies.

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>
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, 21(4), 503-546. <https://doi.org/10.2307/1907921>
- Anderson, E. (2004). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia*, 19(1), 1-24. <https://doi.org/10.1111/j.1527-2001.2004.tb01266.x>
- Ariely, D. (2008). *Predictably irrational*. New York: HarperCollins.
- Ball, S., & Head, B. W. (2020). Behavioural insights teams in practice: nudge missions and methods on trial. *Policy & Politics*, 49(1), 105-120. <https://doi.org/10.1332/030557320X15840777045205>
- Basel, J. S., & Brühl, R. (2013). Rationality and dual process models of reasoning in managerial cognition and decision making. *European Management Journal*, 31(6), 745-754. <https://doi.org/10.1016/j.emj.2013.07.004>
- Behavioural Insights Team. (2012). *Test, learn, adapt: Developing public policy with randomised controlled trials*. London: Cabinet Office. <https://www.bi.team/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials/>
- Behavioural Insights Team. (2016). *Applying behavioural insights to regulated markets*. London: Cabinet Office. <https://www.bi.team/publications/applying-behavioural-insights-to-regulated-markets/>
- Berg, N., & Gigerenzer, G. (2006). Peacemaking among inconsistent rationalities?. In C. Engel & L. Daston (Eds.), *Is there value in inconsistency?* (pp. 421-433). Bader-Baden: Nomos-Verlags-Gesellschaft.
- Bernhardt, A.M., & Reichenspurner, H. (2014). High-risk donors: Extending our criteria in times of organ shortage. *Current Opinion in Organ Transplantation*, 19(5), 494-499. doi: 10.1097/MOT.0000000000000118
- Bernoulli, D. (1954). Specimen theoriae novae de mensura sortis (L. Sommer, Trans.). *Econometria*, 22, 23-36. (Original work published 1738). (pp. 23-36).
- Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, 3(2), 207-220. <https://doi.org/10.1007/s13194-012-0062-x>

Binmore, K. (2008). *Rational decisions*. Princeton: Princeton University Press.

Binmore, K. (2009). 'Fairness as a natural phenomenon', <https://discovery.ucl.ac.uk/id/eprint/14995/1/14995.pdf>, (accessed 20 October 2022).

Boaz, A., Ashby, D., & Young, K. (2002). *Systematic reviews: what have they got to offer evidence-based policy and practice?*. Working Paper 2. London: ESRC UK Centre for Evidence-based Policy and Practice.

Bovens, L. (2009). The ethics of nudge. In T. Grüne-Yanoff & S.O. Hansson (Eds.), *Preference change: Approaches from philosophy, economics and psychology* (pp. 207–219). New York: Springer. <https://doi.org/10.1007/978-90-481-2593-7>

Bradley, R. (2017). *Decision theory with a human face*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/S0266267118000408>

Bristow, D., Carter, L., & Martin, S. (2015). Using evidence to improve policy and practice: The UK What Works Centres. *Contemporary Social Science*, 10(2), 126-137. <https://doi.org/10.1080/21582041.2015.1061688>

Broome, J. (1991). Utility. *Economics & Philosophy*, 7(1), 1-12. <https://doi.org/10.1017/S0266267100000882>

Cadario, R., & Chandon, P. (2020). Which healthy eating nudges work best? A meta-analysis of field experiments. *Marketing Science*, 39(3), 465-486. <https://doi.org/10.1287/mksc.2018.1128>

Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2009). Optimal defaults and active decisions. *The Quarterly Journal of Economics*, 124(4), 1639-1674. <https://doi.org/10.1162/qjec.2009.124.4.1639>

Cartwright, N. (2009). Evidence-based policy: What's to be done about relevance?. *Philosophical Studies*, 143(1), 127-136. <https://doi.org/10.1007/s11098-008-9311-4>

Cartwright, N. (2012). Presidential address: Will this policy work for you? Predicting effectiveness better: How philosophy helps. *Philosophy of Science*, 79(5), 973-989. <https://doi.org/10.1086/668041>

Cartwright, N., Goldfinch, A., & Howick, J. (2009). Evidence-based policy: Where is our theory of evidence?. *Journal of Children's Services*, 4(4), 6-14. <https://doi.org/10.5042/jcs.2010.0017>

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. New York: Oxford University Press. <https://doi.org/10.5042/jcs.2010.0017>

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, 33(2), 339-360. <https://doi.org/10.1007/s11245-013-9220-9>

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58(1), 1-73. [https://doi.org/10.1016/0010-0277\(95\)00664-8](https://doi.org/10.1016/0010-0277(95)00664-8)

Davies, H. T., Nutley, S. M., & Smith, P. C. (2000). What works?: Evidence-based policy and practice in public services. Bristol: Policy Press.

De Cruz, H., & De Smedt, J. (2013). The value of epistemic disagreement in scientific practice. The case of Homo floresiensis. *Studies in History and Philosophy of Science Part A*, 44(2), 169-177. <https://doi.org/10.1016/j.shpsa.2013.02.002>

Dede, O. C. (2019). Behavioral policies and inequities: The case of incentivized smoking cessation policies. *Journal of Economic Methodology*, 26(3), 272-289. <https://doi.org/10.1080/1350178X.2019.1625223>

Dede, O. C. (2021). *Science and values: A philosophical perspective on the justifiability of evidence-based policymaking* [Doctoral thesis]. Erasmus University Rotterdam. <http://hdl.handle.net/1765/137059>

Delgado, J., Molina-Pérez, A., Shaw, D., & Rodríguez-Arias, D. (2019). The role of the family in deceased organ procurement: A guide for clinicians and policymakers. *Transplantation*, 103(5), e112-e118. doi: 10.1097/TP.0000000000002622

Dinner, I., Johnson, E. J., Goldstein, D. G., & Liu, K. (2011). Partitioning default effects: Why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17(4), 332. <https://psycnet.apa.org/doi/10.1037/a0024354>

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, 67(4), 559-579. <https://doi.org/10.1086/392855>

Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese*, 138(3), 453-473. <https://doi.org/10.1023/B:SYNT.0000016451.18182.91>

Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75(4), 643-669. <https://doi.org/10.2307/1884324>

Ensaff, H., Homer, M., Sahota, P., Braybrook, D., Coan, S., & McLeod, H. (2015). Food choice architecture: An intervention in a secondary school and its impact on students' plant-based food choices. *Nutrients*, 7(6), 4426-4437. <https://doi.org/10.3390/nu7064426>

- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European Review of Social Psychology*, 2(1), 83-115. <https://doi.org/10.1080/14792779143000033>
- Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. New York: Oxford University Press.
- Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. In D.J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 62-88). Malden: Blackwell Publisher Ltd. <https://psycnet.apa.org/doi/10.1002/9780470752937>
- Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology*, 6(3), 361-383. <https://doi.org/10.1007/s13164-015-0248-1>
- Gigerenzer, G. (2021). Axiomatic rationality and ecological rationality. *Synthese*, 198(4), 3547-3564. <https://doi.org/10.1007/s11229-019-02296-5>
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1), 107-143. <https://doi.org/10.1111/j.1756-8765.2008.01006.x>
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4), 684–704. <https://psycnet.apa.org/doi/10.1037/0033-295X.102.4.684>
- Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53-96. <https://doi.org/10.1111/j.1539-6053.2008.00033.x>
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109(1), 75. <https://psycnet.apa.org/doi/10.1037/0033-295X.109.1.75>
- Grayot, J. D. (2020). Dual process theories in behavioral economics and neuroeconomics: A critical review. *Review of Philosophy and Psychology*, 11(1), 105-136. <https://doi.org/10.1007/s13164-019-00446-9>
- Grüne-Yanoff, T. (2012). Old wine in new casks: Libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, 38(4), 635-645. <https://doi.org/10.1007/s00355-011-0636-0>

- Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics & Philosophy*, 32(3), 463-483. <https://doi.org/10.1017/S0266267115000425>
- Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory?. *Minds and Machines*, 26(1), 149-183. <https://doi.org/10.1007/s11023-015-9367-9>
- Grüne-Yanoff, T., Marchionni, C., & Feufel, M. A. (2018). Toward a framework for selecting behavioural policies: How to choose between boosts and nudges. *Economics & Philosophy*, 34(2), 243-266. <https://doi.org/10.1017/S0266267118000032>
- Grüne-Yanoff, T. (2007). Bounded rationality. *Philosophy Compass*, 2(3), 534-563. <https://doi.org/10.1111/j.1747-9991.2007.00074.x>
- Gundersen, T. (2018). Scientists as experts: A distinct role?. *Studies in History and Philosophy of Science Part A*, 69, 52-59. <https://doi.org/10.1016/j.shpsa.2018.02.006>
- Gundersen, T. (2020). Value-free yet policy-relevant? The normative views of climate scientists and their bearing on philosophy. *Perspectives on Science*, 28(1), 89-118. https://doi.org/10.1162/posc_a_00334
- Gustafsson, J. E. (2022). *Money-pump arguments* (Elements in Decision Theory and Philosophy). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108754750>
- Hafenbrädl, S., Waeger, D., Marewski, J. N., & Gigerenzer, G. (2016). Applied decision making with fast-and-frugal heuristics. *Journal of Applied Research in Memory and Cognition*, 5(2), 215-231. <https://doi.org/10.1016/j.jarmac.2016.04.011>
- Halpern, D. (2015). *Inside the nudge unit: How small changes can make a big difference*. London: Random House.
- Hammond, P. J. (1998). Subjective expected utility. In S. Barbera, P. Hammond & C. Seidl (Eds.), *Handbook of Utility Theory: Volume 1: Principles* (pp. 213-271).
- Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove?. *European Journal of Risk Regulation*, 7(1), 155-174. <https://doi.org/10.1017/S1867299X00005468>
- Hausman, D. M. (2000). Revealed preference, belief, and game theory. *Economics & Philosophy*, 16(1), 99-115. <https://doi.org/10.1017/S0266267100000158>
- Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy*, 18(1), 123-136. <https://doi.org/10.1111/j.1467-9760.2009.00351.x>
- Head, B. W. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration*, 67(1), 1-11. <https://doi.org/10.1111/j.1467-8500.2007.00564.x>

- Head, B. W. (2010). Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society*, 29(2), 77-94. <https://doi.org/10.1016/j.polsoc.2010.03.001>
- Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12(4), 275-305. [https://doi.org/10.1002/\(SICI\)1099-0771\(199912\)12:4%3C275::AID-BDM323%3E3.0.CO;2-M](https://doi.org/10.1002/(SICI)1099-0771(199912)12:4%3C275::AID-BDM323%3E3.0.CO;2-M)
- Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), 973-986. <https://doi.org/10.1177/1745691617702496>
- Heukelom, F. (2007). *Kahneman and Tversky and the origin of behavioral economics*. Tinbergen Institute Discussion Paper No. 07-003/1. <https://dx.doi.org/10.2139/ssrn.956887>
- Heukelom, F. (2010). Measurement and decision making at the University of Michigan in the 1950s and 1960s. *Journal of the History of the Behavioral Sciences*, 46(2), 189-207. <https://doi.org/10.1002/jhbs.20425>
- Heukelom, F. (2014). *Behavioral economics: A history*. New York: Cambridge University Press. <https://doi.org/10.4000/oeconomia.1498>
- Hoffrage, U., & Reimer, T. (2004). Models of bounded rationality: The approach of fast and frugal heuristics. *Management Review*, 15(4), 437-459.
- Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, 23(1), 1-25. <https://doi.org/10.1080/1350178X.2015.1070527>
- Jachimowicz, J. M., Duncan, S., & Weber, E. U. (2016). Default-switching: The hidden cost of defaults [Working paper]. Retrieved from <http://www.ssrn.com/abstract=2727301>
- Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 3(2), 159-186. <https://doi.org/10.1017/bpp.2018.43>
- Jenny, M. A., Pachur, T., Williams, S. L., Becker, E., & Margraf, J. (2013). Simple rules for detecting depression. *Journal of Applied Research in Memory and Cognition*, 2(3), 149-157. <https://doi.org/10.1016/j.jarmac.2013.06.001>
- Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives?. *Science*, 302(5649), 1338-1339. <https://doi.org/10.1126/science.1091721>
- Johnson, E. J., & Goldstein, D. G. (2004). Defaults and donation decisions. *Transplantation*, 78(12), 1713-1716. doi: 10.1097/01.TP.0000149788.10382.B2

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

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). Cambridge: Cambridge University Press.
<https://psycnet.apa.org/doi/10.1017/CBO9780511808098>

Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition*, 7(4), 409–411. [https://psycnet.apa.org/doi/10.1016/0010-0277\(79\)90024-6](https://psycnet.apa.org/doi/10.1016/0010-0277(79)90024-6)

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In L.C. MacLean & W.T. Ziemba (Eds.), *Handbook of the Fundamentals of Financial Decision Making: Part I* (pp. 99-127). Singapore: World Scientific. <https://doi.org/10.1142/8557>

Kaiser, F., Arnold, O., & Otto, S. (2014). Attitudes and defaults save lives and protect the environment jointly and compensatorily: Understanding the behavioral efficacy of nudges and other structural interventions. *Behavioral Sciences*, 4(3), 202-212.
<https://doi.org/10.3390/bs4030202>

Keller, P. A., Harlam, B., Loewenstein, G., & Volpp, K. G. (2011). Enhanced active choice: A new method to motivate behavior change. *Journal of Consumer Psychology*, 21(4), 376-383.
<https://doi.org/10.1016/j.jcps.2011.06.003>

Khosrowi, D. (2019). Extrapolation of causal effects—hopes, assumptions, and the extrapolator’s circle. *Journal of Economic Methodology*, 26(1), 45-58.
<https://doi.org/10.1080/1350178X.2018.1561078>

Kitcher, P. (2003). *Science, truth, and democracy*. New York: Oxford University Press.

Kitcher, P. (2011). Science in a democratic society. In W.J. Gonzalez (Eds.), *Scientific realism and democratic society* (pp. 95-112). New York: Brill.

Kourany, J. A. (2010). *Philosophy of science after feminism*. New York: Oxford University Press.

Krijnen, J. M., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy*, 3(2), i-18.
<https://doi.org/10.1353/bsp.2017.0010>

Kruglanski, A. W., & Gigerenzer, G. (2011). “Intuitive and deliberate judgments are based on common principles”: Correction to Kruglanski and Gigerenzer (2011). *Psychological Review*, 118(3), 522. <https://psycnet.apa.org/doi/10.1037/a0023709>

- Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication: Graphical and analog tools. *Annals of the New York Academy of Sciences*, 1128(1), 18-28. <https://doi.org/10.1196/annals.1399.004>
- Lacey, H. (1999). Scientific understanding and the control of nature. *Science & Education*, 8(1), 13-35. <https://doi.org/10.1023/A:1008678413285>
- Lacey, H. (2002). The ways in which the sciences are and are not value free. In P. Gärdenfors, J. Woléński & K. Kijani-Placek (Eds.), *In the scope of logic, methodology and philosophy of science* (pp. 519-532). Dordrecht: Springer. <https://doi.org/10.1007/978-94-017-0475-5>
- Lindman, H. R. (1971). Inconsistent preferences among gambles. *Journal of Experimental Psychology*, 89(2), 390. <https://psycnet.apa.org/doi/10.1037/h0031208>
- Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton: Princeton University Press. <https://doi.org/10.2307/j.ctvx5wbfz>
- Longino, H. E. (1996). Cognitive and non-cognitive values in science: Rethinking the dichotomy. In L.H. Nelson & J. Nelson (Eds.), *Feminism, science, and the philosophy of science* (pp. 39-58). Dordrecht: Springer.
- Lynge, E. (1998). Mammography screening for breast cancer in Copenhagen April 1991-March 1997. Mammography Screening Evaluation Group. *APMIS*, 83(Suppl), 1-44. PMID: 9850674
- Maher, P. (1993). *Betting on theories*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/S026626710000482X>
- Malecka, M. (2020). The normative decision theory in economics: A philosophy of science perspective. The case of the expected utility theory. *Journal of Economic Methodology*, 27(1), 36-50. <https://doi.org/10.1080/1350178X.2019.1640891>
- Mandler, M. (2001). A difficult choice in preference theory: Rationality implies completeness or transitivity but not both. In E. Millgram (Eds.), *Varieties of practical reasoning* (pp. 373-402). Cambridge: MIT Press.
- Marchionni, C., & Reijula, S. (2019). What is mechanistic evidence, and why do we need it for evidence-based policy?. *Studies in History and Philosophy of Science Part A*, 73, 54-63. <https://doi.org/10.1016/j.shpsa.2018.08.003>
- Marck, C. H., Jansen, N., Neate, S. L., Dwyer, B. M., van Haren, F., Weimar, W., & IJzermans, J. N. (2014). Comparing organ donation practices and rates between Australia and the Netherlands to detect best practices and areas for improvement. *Organs, Tissues & Cells*, 17, 175-83.

- Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing*, *11*(2), 103-121. <https://doi.org/10.1007/s10339-009-0337-0>
- Markowitz, H. (1952). The utility of wealth. *Journal of political Economy*, *60*(2), 151-158.
- Marlais, M., Pankhurst, L., Hudson, A., Sharif, K., & Marks, S. D. (2017). UK National Registry Study of kidney donation after circulatory death for pediatric recipients. *Transplantation*, *101*(6), 1177-1181. doi: 10.1097/TP.0000000000001264
- McMullin, E. (1982). Values in science. *PSA: Proceedings of the biennial meeting of the philosophy of science association*, *1982*(2), 2-28. <https://doi.org/10.1086/psaprocbienmeetp.1982.2.192409>
- Mikkelsen, B. E., Sudzina, F., Ørnbo, L. E., & Tvedebrink, T. D. O. (2021). Does visibility matter?—A simple nudge reduces the purchase of sugar sweetened beverages in canteen drink coolers. *Food Quality and Preference*, *92*, 104190. <https://doi.org/10.1016/j.foodqual.2021.104190>
- Mitchell, S. (2004). The prescribed and proscribed values in science policy. In P. Machamer & G. Wolters (Eds.), *Science, values, and objectivity* (pp. 245-55). Pittsburgh: University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt5vkg7t>
- Mongin, P. (1998). Expected utility theory. In J.B. Davis, D.W. Hands & U. Mäki (Eds.), *The Handbook of Economic Methodology* (pp. 171-179). Cheltenham: Edward Elgar
- Mongin, P., & Baccelli, J. (2021). Expected utility theory, Jeffrey's decision theory, and the paradoxes. *Synthese*, *199*(1), 695-713. <https://doi.org/10.1007/s11229-020-02691-3>
- Mongin, P., & Cozic, M. (2018). Rethinking nudge: Not one but three concepts. *Behavioural Public Policy*, *2*(1), 107-124. <https://doi.org/10.1017/bpp.2016.16>
- Moscato, I. (2016). Retrospectives: How economists came to accept expected utility theory: The case of Samuelson and Savage. *Journal of Economic Perspectives*, *30*(2), 219-36. doi: 10.1257/jep.30.2.219
- Nagatsu, M. (2015). Behavioral economics, history of. *International Encyclopedia of the Social & Behavioral Sciences*, *2*(2), 443-449. <http://dx.doi.org/10.1016/B978-0-08-097086-8.03053-1>
- Newman, J. (2017). Deconstructing the debate over evidence-based policy. *Critical Policy Studies*, *11*(2), 211-226. <https://doi.org/10.1080/19460171.2016.1224724>
- Nisa, C. F., Bélanger, J. J., Schumpe, B. M., & Faller, D. G. (2019). Meta-analysis of randomized controlled trials testing behavioural interventions to promote household action on climate change. *Nature Communications*, *10*(1), 1-13. <https://doi.org/10.1038/s41467-019-12457-2>

OECD (Organisation for Economic Co-operation and Development). (2017). *Behavioural insights and public policy: Insights from around the world*. Paris: OECD Publishing.
<https://doi.org/10.1787/9789264270480-en>

Plott, C. R. (1996). Comments on: Daniel Kahneman, "New challenges to the rationality assumption". In K. Arrow, E. Colombatto, M. Perlman & C. Schmidt (Eds.), *The rational foundations of economic behavior*. London: MacMillan.

Rebonato, R. (2012). *Taking liberties: A critical examination of libertarian paternalism*. New York: Palgrave Macmillan.

Rebonato, R. (2014). A critical assessment of libertarian paternalism. *Journal of Consumer Policy*, 37(3), 357-396. <https://doi.org/10.1007/s10603-014-9265-1>

Rostain, T. (2000). Educating homo economicus: Cautionary notes on the new behavioral law and economics movement. *Law and Society Review*, 34(4), 973-1006.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science*, 20(1), 1-6. <https://doi.org/10.1086/287231>

Samuels, R., Stich, S., & Bishop, M. (2012). Ending the rationality wars. In S. Stich (Eds.), *Collected papers, volume 2: Knowledge, rationality, and morality, 1978-2010* (pp. 191-223). New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199733477.001.0001>

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policymaking. *Public Administration*, 80(1), 1-22. <https://doi.org/10.1111/1467-9299.00292>

Savage, L. J. (1972). *The foundations of statistics*. New York: Dover Publications.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, 130(3), 380-400.
<https://psycnet.apa.org/doi/10.1037/0096-3445.130.3.380>

Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys*, 16(5), 621-655.
<https://doi.org/10.1111/1467-6419.00184>

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 6(4), 317-344.

Sent, E.-M. (2004). Behavioral Economics: How psychology made its (limited) way back into economics. *History of Political Economy*, 36(4), 735-760.

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), 1-14. <https://doi.org/10.1186/s12916-014-0131-4>

- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99-118. <https://doi.org/10.2307/1884852>
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129-138. <https://psycnet.apa.org/doi/10.1037/h0042769>
- Simon, H. A. (1976). From substantive to procedural rationality. In T.J. Kastelein, S.K. Kuipers, W.A. Nijenhuis & G.R. Wagenaar (Eds.), *25 years of economic theory* (pp. 65-86). Boston: Springer. <https://doi.org/10.1007/978-1-4613-4367-7>
- Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology*, 78(3, Pt.2), 1-18. <https://psycnet.apa.org/doi/10.1037/h0026468>
- Smith, N. C., Goldstein, D. G., & Johnson, E. J. (2013). Choice without awareness: Ethical and policy implications of defaults. *Journal of Public Policy & Marketing*, 32(2), 159-172. <https://doi.org/10.1509/jppm.10.114>
- Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. New York: Psychology Press. <https://doi.org/10.4324/9781410603432>
- Steel, D. (2010). Epistemic values and the argument from inductive risk. *Philosophy of Science*, 77(1), 14-34. <https://doi.org/10.1086/650206>
- Steel, D. (2016). Climate change and second-order uncertainty: Defending a generalized, normative, and structural argument from inductive risk. *Perspectives on Science*, 24(6), 696-721. https://doi.org/10.1162/POSC_a_00229
- Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science*. New York: Oxford University Press.
- Stevens, J., Tumin, D., Shaffer, K. L., Bickman, L., Hoagwood, K. E., & Hayes Jr, D. (2019). Are there missed opportunities to maximize organ donation registrations? An examination of driver's license applications across the United States. *Progress in Transplantation*, 29(2), 173-178. <https://doi.org/10.1177/1526924819835832>
- Sugden, R. (2022). Debiasing or regularisation? Two interpretations of the concept of 'true preference' in behavioural economics. *Theory and Decision*, 92(3), 765-784. <https://doi.org/10.1007/s11238-022-09876-x>
- Sunstein, C. R. (2018). Misconceptions about nudges. *Journal of Behavioral Economics for Policy*, 2(1), 61-67.

- Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1), S164-S187. <https://doi.org/10.1086/380085>
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, 93(2), 175-179. doi: 10.1257/000282803321947001
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. London: Penguin.
- Timmons, S., Robertson, D. A., & Lunn, P. D. (2022). Combining nudges and boosts to increase precautionary saving: A large-scale field experiment. ESRI Working Paper No. 722. Economic and Social Research Institute.
- Todd, P. M., & Brighton, H. (2016). Building the theory of ecological rationality. *Minds and Machines*, 26(1), 9-30. <https://doi.org/10.1007/s11023-015-9371-0>
- Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23(5), 727-741. <https://doi.org/10.1017/S0140525X00003447>
- Todd, P. M., & Gigerenzer, G. E. (2012). *Ecological rationality: Intelligence in the world*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195315448.001.0001
- Truijens, D. (2022). Coherence between theory and policy in Nudge and Boost: Is it relevant for evidence-based policymaking?. *Rationality and Society*, 34(3), 368-394. <https://doi.org/10.1177/10434631221093741>
- Truijens, D., & van Exel, J. (2019). Views on deceased organ donation in the Netherlands: A q-methodology study. *PLoS ONE*, 14(5), e0216479. <https://doi.org/10.1371/journal.pone.0216479>
- Tversky, A., & Kahneman, D. (1985). The framing of decisions and the psychology of choice. In G. Wright (Eds.), *Behavioral decision making* (pp. 25-41). Boston: Springer. <https://doi.org/10.1007/978-1-4613-2391-4>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- Tversky, A., & Kahneman, D. (1986). The framing of decisions and the evaluation of prospects. *Studies in Logic and the Foundations of Mathematics*, 114, 503-520. [https://doi.org/10.1016/S0049-237X\(09\)70710-4](https://doi.org/10.1016/S0049-237X(09)70710-4)

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. In B. Karpak & S. Zionts (Eds.), *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81-126). Berlin: Springer. <https://doi.org/10.1007/978-3-642-74919-3>

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323. <https://doi.org/10.1007/BF00122574>

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323. <https://doi.org/10.1007/BF00122574>

Van Gestel, L. C., Kroese, F. M., & De Ridder, D. T. D. (2018). Nudging at the checkout counter—A longitudinal study of the effect of a food repositioning nudge on healthy food choice. *Psychology & Health*, 33(6), 800-809. <https://doi.org/10.1080/08870446.2017.1416116>

Vlaev, I., King, D., Dolan, P., & Darzi, A. (2016). The theory and practice of “nudging”: changing health behaviors. *Public Administration Review*, 76(4), 550-561. <https://doi.org/10.1111/puar.12564>

Von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Vranas, P. B. (2000). Gigerenzer's normative critique of Kahneman and Tversky. *Cognition*, 76(3), 179-193. [https://doi.org/10.1016/S0010-0277\(99\)00084-0](https://doi.org/10.1016/S0010-0277(99)00084-0)

Wall, S. P., Plunkett, C., & Caplan, A. (2015). A potential solution to the shortage of solid organs for transplantation. *Jama*, 313(23), 2321-2322. doi:10.1001/jama.2015.5328

Wilson, A. L., Buckley, E., Buckley, J. D., & Bogomolova, S. (2016). Nudging healthier food and beverage choices through salience and priming. Evidence from a systematic review. *Food Quality and Preference*, 51, 47-64. <https://doi.org/10.1016/j.foodqual.2016.02.009>

Wilson, D. K., Kaplan, R. M., & Schneiderman, L. J. (1987). Framing of decisions and selections of alternatives in health care. *Social Behaviour*, 2(1), 51-59.

Part 1.

BEHAVIOR CHANGE

Chapter 3.

Views on deceased organ donation in the Netherlands: A q-methodology study

D.C. Truijens
N.J.A. van Exel*

PLoS ONE 2019; 14(5): e0216479.

* Both authors have equally contributed to this research. D.C. Truijens has set up the experiment and conducted the interviews. N.J.A. van Exel has executed the statistical analysis. The writing of this article was shared equally.

3. Views on deceased organ donation in the Netherlands

Abstract

In many countries, such as the US, Germany, France, and the Netherlands, governments are dealing with a great shortage of organ donors. Even though people generally show positive attitudes towards organ donation, they often do not actually register as organ donors themselves. This study's objective was to explore prevailing viewpoints among the Dutch population on deceased organ donation and the relation between aspects of the viewpoints potentially influencing the decision to register as an organ donor. Although substantive research about attitudes on organ donation has been conducted, this is the first study investigating people's viewpoints focusing on the *relation* between beliefs, tastes, preferences, motives, goals and other constituents underlying people's viewpoints on organ donation, such as the role of the media and public policies. This Q-methodology study revealed four viewpoints: "not donating your organs is a waste", "it does not go with my religion", "my family should decide"; and "it's a good deed, but I'm doubtful". These viewpoints convey information on potential reasons for the gap between people's favorable attitudes towards organ donation and the low number of actual registrations, and opportunities for policy makers to address certain target groups more adequately.

3.1 Introduction

In many countries, such as the US, Germany, France, and the Netherlands, governments are dealing with a great shortage of organ donors (e.g., Bernhardt and Reichensperner 2014, Marck et al. 2014, Marlais et al. 2017, Wall, Plunkett and Caplan 2015). Even though people generally show positive attitudes towards organ donation, they often do not actually register as organ donors themselves (Feeley 2007, Levitt et al. 2011, Manninen and Evans 1985, Radecki and Jaccard 1997, Rodrigue, Cornell and Howard 2006, Siegel et al. 2014, Stevens 2013, Quick et al. 2016). In order to create greater awareness of organ donor scarcity and ultimately recruit more potential organ donors, public policy makers have taken up a wide range of actions, including the development of extensive media campaigns (e.g., Feeley and Moon 2009, Morgan 2009, Rady and Verheijde 2014). In this context, public policy increasingly relies on behavioral insights from economics and psychology, such as research which shows that people tend to stick with default positions and are therefore more likely to become an organ donor under an opt-out system (e.g., Abadie and Gay 2006, Dinner et al. 2011, Johnson and Goldstein 2003, Johnson and Goldstein 2004, Kennedy et al. 1998, Orentlicher 2008, Spetherd, O'Carroll and Ferguson 2014, Thaler and Sunstein 2008, Smith, Goldstein and Johnson 2013). For instance, in the Netherlands recently a law passed which will change the organ donation registration system from an opt-in to an opt-out system. Positive effects of these type of behavioral public policies are often explained by weakness of will or by people's alleged indifference about the choice at issue (Dinner et al. 2011, Smith, Goldstein and Johnson 2013). Yet, these explanations are considered controversial (e.g., Smith Goldstein and Johnson 2013), and the question of how people think about organ donation and the choice to register as organ donors, including the role of default positions, is still a matter of debate (e.g., Jachimowicz, Duncan and Weber 2016).

In order to develop adequate and effective public policies to promote organ donor registration, it is essential to gain insight in people's viewpoints on organ donation. Understanding the existing viewpoints on this issue is important to be able to identify target groups for policy, and how to approach them adequately and effectively. Some citizens will have made up their mind about registering as an organ donor, others might still be in doubt, for a variety of reasons. These various groups require different approaches in terms of raising awareness and knowledge, and incentivizing registration. Moreover, it may inform about potential drawbacks of policies for particular subgroups that need consideration. For instance, implementing an opt-out policy might disadvantage people who are against organ donation but for whom information on the registration process and instructions for opting out are less accessible. In this context, several studies have investigated people's attitudes towards organ donation and their willingness to register as an organ donor (e.g., Hübner, Mohs and Petersen 2014, Morgan et al. 2013, Spetherd and O'Carroll 2014). Furthermore, substantive research has been done on sociodemographic differences in attitudes, knowledge and behaviour towards organ donation (e.g., Falomir-Pichastor, Berent and Pereira 2013, Güden, Cetinkaya and Nacar 2013, Hyde, Knowles and White 2013, Mohamed and Guella 2013, Morgan and Miller 2002, Murray et al. 2013, Zimmerman et al. 2006). However, little research has been conducted into how people's beliefs, preferences, goals, and other elements such as the perceived role of the media and public policy are related and together constitute different viewpoints on organ donation that exist among the public. As a consequence, there is insufficient knowledge of what characterizes and distinguishes people's viewpoints, and how this could inform the development of adequate and effective policies to promote organ donation. The aim of this research was to contribute to the literature by exploring more in detail the prevailing viewpoints about deceased organ donation among the public in the Netherlands, and their implications for the decision to register as an organ donor. Deceased organ donation refers to organ donation after death, thereby excluding "living organ donation" of, for instance, kidneys.

3.2 Methods

3.2.1 Q-methodology

We used Q-methodology to explore the viewpoints about organ donation among the public in the Netherlands. Q-methodology combines qualitative and quantitative methods to discover what constitutes the view of individual respondents about a value-laden topic, identify shared viewpoints among respondents, and the consensus and differences between these viewpoints (Shemmings 2006, Watts and Stenner 2012). In this context, viewpoints are defined as an attitude how one thinks of organ donation, constituted by a coherent set of individual beliefs, preferences and goals, the perceived role of the media and public policy, and other elements potentially influencing how one perceives the decision to register as an organ donor (for more detailed discussions of the definition of viewpoints in the context of Q-methodology studies (e.g., Akhtar-Danesh et al. 2007, Baker, Thompson and Mannion 2006, Bryant et al. 2017). Because it commonly relies on purposive sampling, Q-methodology is useful to explore the variety of viewpoints that exists within a population but does not provide information about the proportion of the population that holds a particular viewpoint (Kitzinger 1987). In a Q study, respondents are asked to rank a comprehensive set of statements concerning the topic at issue and to explain their ranking of the statements in a follow-up interview. The shared viewpoints are identified using by-person factor analysis, which

searches for correlations between respondents' answers across a sample of variables (Brown 1980). The underlying assumptions are that within a certain population only a limited number of viewpoints exist, thereby paving the way for an analysis of types of viewpoints rather than assessing each individual's viewpoint by itself, and that respondents who rank the statements in a similar way share a similar viewpoint on the topic (Brown 1995). In this study we apply common techniques in Q-methodology for the design of the study, and collection and analysis of the data. Readers interested in more detail about the background of the method and common/ best practices in applications of the method are referred to Watts and Stenner (Watts and Stenner 2012) or McKeown and Thomas (McKeown and Thomas 2013). Q-methodology has been applied before in the domains of health promotion, education and behavior (e.g., Akhtar-Danesh et al. 2007, Patty et al. 2017, Vermaire et al. 2010, Tielen et al. 2008, Van Exel, De Graaf and Brouwer 2006, Cross 2004, Risdon et al. 2003).

3.2.2 Development of the statement set

For this study, a set of 43 statements was constructed, containing a large variety of aspects relating to the decision to register for deceased organ donation. In order to enable all respondents to express their view it is important that the statement set is comprehensive and balanced, broadly representative of the issues of debate on this topic in the literature and among the target population. For this purpose, we first identified potentially relevant aspects to cover from common theoretical models about health intentions and behavior like the Health Belief Model (Rosenstock 1974) and the Theory of Planned Behaviour (Ajzen 1991). This concerned perceived severity, susceptibility, barriers and benefits, self-efficacy / perceived behavioral control, cues to action and subjective norms. In addition, based on the recent literature on determinants of the decision to register as an organ donor (e.g., Rady and Verheijde 2014, Davison and Jhangri 2014), a "moral beliefs" component was added.

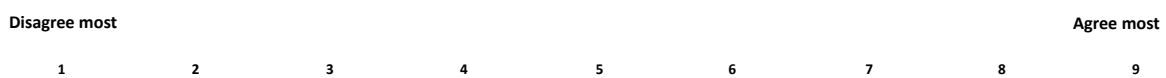
It should be stressed that the aim of using these theoretical models was to make sure that the statement set is comprehensive and captures all relevant aspects potentially relevant for people's viewpoints on organ donation, not to function as a model by which the causal relations of these aspects are explained.

Next, 350 statements on organ donation were extracted from academic literature, newspaper articles, radio and television documentaries, discussions on social media, and related information websites. Academic literature published between 2003 and 2017 was searched through Google Scholar and PubMed using the keywords "organ donation", "viewpoints on organ donation", "attitudes towards organ donation", "opinions on organ donation", "organ donation default positions", "organ donation opt-in opt-out systems". A total of 350 statements was retrieved, which were structured according to the components of the theoretical models to ensure these were all covered. Redundant, unclear and irrelevant statements were deleted, leaving an initial set of 47 statements for pilot-testing.

In order to check whether the statement set was comprehensive and intelligible, all other research materials were clear, and the whole approach feasible and sufficiently appealing to participants, two pilot studies were conducted. Respondents varying in age, gender, education level, ethnicity and

religiosity were presented with the draft research material and asked to critically review the information letter, instructions and statement set. Based on a first pilot study in which five respondents took part, several changes were made to the statement set, including the rephrasing of several statements, deleting three statements that were considered irrelevant by the respondents and combining two statements that were regarded as too similar. Next, another five respondents and two researchers experienced with Q-methodology were asked to review the adjusted study materials. As a result, one more statement was deleted, four statements were rephrased and one statement was added, resulting in a set of 43 statements for the main study. Moreover, because participants in both rounds of the pilot indicated that the sorting grid did not provide sufficient opportunity to rank statements in the least and most agree categories, the number of spaces for placing statements in the two outer columns of the grid was expanded from two to three (Fig 1).

Figure 1.



3.2.3 Data collection

In the main study, a total of 30 respondents were interviewed (Table 1). Provided they are well-selected, this is generally considered an adequate sample size for exploring the views on a topic (Watts and Stenner 2012). Respondents were recruited via an external agency and sampled purposively to represent demographic and cultural diversity. Based on the literature, the characteristics gender, age, education level and religion were used for sampling.

Table 1.

Characteristic	%
Gender	
Male	16
Female	14
Age	
18–35	11
35–70	19
Highest finished education	
Lower education (primary school, high school, vocational training)	18
Higher education (BA or MA / applied science or university level)	12
Religion & Philosophy of Life	
Atheism	9
Christianity	7
Judaism	3
Islam	6
Buddhism	3
Hinduism	2

Note. From “Views on deceased organ donation in the Netherlands: A q-methodology study”, by D. Truijens & J. van Exel, 2019, *PLoS ONE*, 14(5), e0216479. CC-BY-NC.

At the start of the interview, respondents received an information letter informing them about the aim of the study, the content of the interview, and the processing and analysis of the data. Next, respondents received an informed consent form and were instructed that they could stop the interview at any time, for any reason. Respondents were required to provide informed consent and to finish the interview to be included in the study.

After the consent form was signed, the respondent received the set of 43 statements printed on cards in a random order. In order to ensure a correct understanding of the statements, the researcher would first read each statement out aloud after which the respondent would read the card herself, briefly comment and then place the card on one of three piles, for statements they agreed with, disagreed with, or felt neutral about. After all statements were placed on one of the three piles, respondents were instructed to read the statements they had placed in the ‘agree with’ pile once again and to rank them according to the level of agreement on the right side of the sorting grid (Fig 1). After that, they ranked the statements they had placed in the ‘disagreed with’ pile on the left side of the grid, and those they had placed in the ‘neutral’ pile in the spaces left in the middle of the grid.

After completing the ranking exercise, respondents were asked to explain the placement of certain statements on the sorting grid; all respondents elaborated on the statements placed in the two outer columns of the grid, and depending on time, on randomly chosen other statements. The interview ended with a short questionnaire consisting of closed questions about socio-demographic characteristics and whether they had registered their preference for organ donation or intended to do so.

The interviews took approximately 45 minutes and were conducted on Erasmus University campus. Respondents received 25 Euros in cash as financial compensation for their time. The study protocol was approved by the Medical Ethics Review Board of the Erasmus Medical Centre (reference number MEC-2017-221).

3.2.4. Data analysis

The 30 rankings of statements were analyzed using by-person factor analysis (i.e., centroid factor extraction followed by varimax rotation) in the dedicated software package PQMethod (Schmolck and Atkinson 2002). These common techniques for Q-methodology (Watts and Stenner 2012) identify groups of respondents that have ranked the 43 statements in a similar way. Inspection of statistical information (i.e., explained variance and number of defining variables per factor) and the qualitative materials collected during the interviews resulted in the choice for a four-factor solution. For each of these factors a composite ranking of the statements was computed based on the rankings significantly associated with the factor and their factor loadings as weight. Such composite ranking represents how a person perfectly correlated with the factor would have ranked the 34 statements. The four factors were first interpreted based on this composite ranking of the statements per factor, considering in particular the characterizing statements (i.e., those ranked in the two outer columns at each side of the composite ranking) and the distinguishing statements (i.e., those ranked statistically significantly different in a factor as compared to all other factors) of each factor. Subsequently, the qualitative materials from the interviews of respondents associated significantly with a factor ($p < .05$) were used to refine the check and interpretation of the factors. A number of citations from respondents were used to illustrate the interpretations.

3.3 Results

The final sample consisted of 30 respondents. Table 1 shows the characteristics of the sample. The four factors explained 47% of the variance in the rankings of statements, and 28 respondents were significantly associated with one of the factors ($p < .05$). Factor correlations ranged from 0.15 for factors 3 and 4 to 0.48 for factors 1 and 2. Table 2 presents the composite rankings of the statements for each of the four factors.

Table 2.

Statement	Viewpoint			
	Not donating my organs would be a waste	It does not go with my religion	My family should decide	It is a good deed, but I'm doubtful
1. 1. I might need a donor organ myself one day	+3	+3	0	+2
2. 2. I expect that, if really necessary, there always will be a donor organ available	-1	0	-1	-1
3. 3. I expect that people waiting for a donor organ may contract health damage	+3	+2	+4	+3
4. 4. People in need of a donor organ are not able to function normally	0	+2	+1	+2
5. 5. Information in the media helps me to determine my preference for registering as an organ donor	+1	0	0	0
6. 6. By donating my organs I can pass on the gift of life to someone else	+4	+4	+3	+1
7. 7. By donating my organs I add meaning to my life	0	-1	-1	+4*
8. 8. By donating my organs I can reduce the suffering of the recipients	+4	+1	+3	+3
9. 9. I like the idea that if I donate my organs part of me will live on after I die	-1	0	-2	0
10. 10. Donating my organs makes me proud	0	+1	+1	+3
11. 11. I think it would be a waste not to donate my organs	+4*	-1	0	+1
12. 12. I would mind donating organs that are visible to my loved ones at my farewell	0*	+3	+2	+2
13. 13. I need my organs for the afterlife	-3	-4	-2	-3
14. 14. My family would not have any problems with me donating my organs	+2*	-3	0	-1
15. 15. I am afraid that when I have registered as an organ donor, doctors may remove my organs while I am still alive	-3	+1*	-3	-4
16. 16. I am afraid that when I have registered as an organ donor, doctors may stop my treatment earlier	-2	-1	-1	-1
17. 17. I am afraid that if I donate my organs, they will end up on the black market	-4	-4	-3	+2*
18. 18. I dislike the idea that if I donate my organs, others will see my naked body	-4	-1*	-4	+1*
19. 19. I would donate some of my organs, but not all	-2*	+1	+1	+1
20. 20. I talk with my friends about organ donation	0	+2	-2*	-4*
21. 21. I find the choice whether to donate my organs difficult	-2	0	+3*	-1
22. 22. Registering your preference for organ donation takes a lot of time and energy	-2	0*	-3	-3
23. 23. I prefer not to think about organ donation	-1	-3	0*	-4
24. 24. I'm afraid that if I donate my organs, someone is kept alive who does not live a good life	-2*	0	-1	0
25. 25. There is sufficient information available about organ donation	+3	-2*	+1	+1
26. 26. I have sufficient knowledge about organ donation	+1*	-3	0	-2
27. 27. I know how to register and unregister as organ donor	+2*	-4	+3*	-1
28. 28. Whenever I'm sure about my preference for donating my organs, I will register it	+3	-3*	+2	+3
29. 29. Organ donation should be compensated financially	-3	-2	-1	-2
30. 30. The family should have the last word about organ donation	-4	-1	+4	+4
31. 31. People who have registered as organ donors should be given priority when they need a donor organ themselves	+1	+1	+1	-3*
32. 32. By donating your organs you do a good deed	+2	+4	+4	+4
33. 33. If you would like to receive a donor organ might you need one, you should also be willing to donate your organs	+1	+3	+2	-2*
34. 34. If you are not registered as organ donor, you should also not be eligible for a donor organ yourself	-1	+1	-4	-3
35. 35. Organ donation is a moral duty	+1	-2	-2	+1
36. 36. Organ donation does not go with my religious beliefs	0	+4*	-4*	0
37. 37. Removing organs from people who are brain dead but are kept alive by machines, is murder	-1	+3	+1	-2
38. 38. One should leave the dead in peace	-3	0	-3	-2
39. 39. The government should not interfere with organ donation, it is an individual matter	0	+2	+2	0
40. 40. People should receive a reward as a nudge to register as organ donor	-1	-2	0	0
41. 41. My loved ones find it important that I register as organ donor	+1	-2	-2	0
42. 42. I talk with my family about organ donation	+2*	-1	-1	-1
43. 43. People who have not registered their preference for organ donation should receive a reminder	+2	+2	+2	+2

Note. From “Views on deceased organ donation in the Netherlands: A q-methodology study”, by D. Truijens & J. van Exel, 2019, *PLoS ONE*, 14(5), p. 6. CC-BY-NC.

Viewpoint 1 – Not donating my organs would be a waste

This can be considered a pro-organ donation perspective, as in this viewpoint the perceived benefits of becoming an organ donor clearly outweigh the perceived barriers. People with this view sympathize with the suffering of people waiting for a donor organ (statement 3 receives factor score +3: st.8, +3), realize they might need an organ someday themselves (st.1, +3) and emphasize benefits of organ donation such as diminishing the suffering of other people (st.8, +4) and being able to pass on life to others (st.6, +4). Distinguishing for this view is the notion that not donating your organs is considered a waste (st.11, +4) as they are of no use to yourself after death (st.13, -3; st.38, -3): *“I really don’t see why you would choose not to donate your organs. If you cannot use your organs anymore yourself, why wouldn’t you help someone else?”*. In line with this, no distinction is made between which organs to donate or not (st.19, -2): *“You no longer need any of them yourself, so as far as I am concerned they can take them all”*; or to whom (st.24, -2).

People with this view see themselves as efficacious. Finding the relevant information is considered as relatively easy (st.22, -2) and they feel they have sufficient knowledge about organ donation (st.25, +1; st.26, +3) and about how to register or unregister as organ donor (st.27, +2): *“Sure, it takes some time, but you search on google, fill out a short form and there you go”*. They discussed the matter with their family (st.42, +2), who do not object to them being organ donor (st.14, +2). More than in the other views, they agreed that their family members encouraged them to register as an organ donor (st.41, +1). Still, they strongly disagreed that family members should have the last say on whether their organs eventually will be donated (st.30, -4), even if family members would object to the donation.

Trust in the healthcare system is widespread within this group. Respondents expected that medical professionals will treat their dead body and its organs with respect (st.15, -3, st.16, -2), and the possibility that the donated organs end up on the black market is considered negligible (st.17, -4). All in all, the choice to register as organ donor was not difficult for them (st.21, -2). People with this view either already had registered as organ donor or stated their intention to do so (st.28, +3), and do not see the need for incentivizing the decision to register (st.29, -3; st.40, -1).

Based on this description, we labelled this viewpoint “Not donating my organs would be a waste”. Thirteen respondents were statistically significantly associated with this viewpoint.

Viewpoint 2 – It does not go with my religion

People with this view strongly agree that organ donation does not match their religious beliefs (st.36, +4). This is not because they expect they will need their organs in the afterlife (st.13, -4), as they believe body and soul will be separated after death. But the body is borrowed and should be returned, and thus be preserved. At the same time, they strongly agreed with the statement “By donating your organs you do a good deed” (st.32, +4) and often even presented a religious reason for this. According to them, people receive the gift of life from God and by organ donation you could pass on this gift to someone else (st.6, +4). Moreover, one day you might need an organ yourself (st.1, +3), and you should be willing to give if you wish to receive (st.33, +3). People holding this view therefore see organ donation as a “paradox” or “personal challenge” that concerns them (st.23, -3); as one of the respondents explained: *“Giving in general is considered a good*

deed and I think it is very important to give. So, for myself, I would really like to become an organ donor. But my religion forbids me to become an organ donor, because you should consider your body to be a loan from God and return it to Him. This is difficult for me, because I think giving and helping other people is an important part of my religion and of who I am, but I have to respect God's will, even though I do not understand it'.

Organ donation is not a common topic of family conversation (st.42, -1). Family members are said to hold a similar religious perspective on this topic and would therefore not support them registering as organ donor (st.14, -3; st.41, -2). However, while they value the opinion of their family highly, the family should not have the last say in the decision to donate their organs (st.30, -1); it is considered to be an individual matter, for which one is personally accountable when faced with God. Notwithstanding, more than those with other views, this group does discuss the topic of organ donation with friends (st.20, +2).

People defining this view say they view would not officially register their preference for becoming organ donor or not (st.28, -3) because under the current opt-in system the only reason to register would be to become organ donor; and that's not what they want, after all. In line with this, they were neutral regarding how difficult it is to register as an organ donor (st.21, 0). They have little knowledge and information about the registration procedure for organ donation or the transplantation process (st.26, -3; st.25,-2), and stated not to know how to register (st.27, -4); the whole procedure is something *"that you just don't really think about"*.

Based on this description, we labelled viewpoint two "It does not go with my religion". Two respondents were statistically significantly associated with this factor.

Viewpoint 3 – My family should decide

People with this view consider the choice to register as an organ donor as very difficult (st.21, +3), which can best be characterized by a dilemma. On the one hand, they see various benefits of organ donation. For instance, they strongly agreed with the statement that organ donation is a good deed (st.32, +4) that would prevent further suffering or perhaps even cure people with severe health problems (st.3, +4; st.6, +3; st.8, +3). On the other hand, respondents argue that the potential suffering of family members constitutes a considerable barrier for them to donate their organs, even though they are not sure whether their family members would actually have problems with them becoming an organ donor (st.42, -1; st.41, -2; st.14, 0). People defining this view often emphasized the rapidity of transplantation and expressed a fear of adding extra trauma to family members if their dead body would be taken away so shortly after death. In addition, respondents viewed the incompleteness of the body after the transplantation as difficult for family members, especially in cases when this would be visible (ref 12, +2). One respondent explained: *"Personally, I find the decision to become an organ donor very difficult. My children are the main reason. I think the moment to say goodbye must be very difficult and terrible for them, especially if you know your mother's body is missing parts. The medical staff also has to take the body away within 24 hours after death in order for the transplantation to be able to succeed. I think this time constraint is especially difficult for the family of the deceased, not being able to stay with the deceased so soon after she passed away must be very unpleasant"*. Since according to this group, the impact of organ donation is on the family of the deceased rather than on the deceased herself, they strongly agreed with the statement that family members should have the final say in deciding

whether the organs of the deceased will be donated (st.30, +4). As this suggests the decision to donate is made post mortem by their family, not being registered as organ donor should not have consequences for their eligibility to receive a donor organ, if needed (st.34, -4).

People with this view claim to know well how to register or unregister as organ donor (st.27, +3), do not think it is too much of an effort (st.22, -3) or conflicts with their philosophy of life (st.36, -4), and do not see significant problems with donating (st.15, -3; st.16, -1; st.17, -3; st.18, -4; st.38, -3). However, they are quite doubtful about whether they have sufficient knowledge on organ donation (st.26, 0) or sufficient information is available to them (st.25, 1).

Based on this description, we labelled viewpoint three “My family should decide”. A total of 8 respondents were statistically significantly associated with this factor.

Viewpoint 4 - It is a good deed, but I’m doubtful

Whereas respondents holding the other viewpoints repeatedly stated that “*organ donation is something that you do for others, not for yourself, to give yourself a good feeling*”, these people see donating your organs as a good deed (st.32, +4) and they would feel proud of themselves (st.10, +3) for doing something that could impact another person’s life positively (st.8, +3; st.3, +3). In this way, registering as organ donor would give more meaning to their own lives (st.7, +4). As one of the respondents defining this view explained: “*registering as an organ donor shows that you value life, not only your own but also the life of others*”. They argued that by registering as organ donor one would endorse the conviction that life is worth living.

Even though people with this view seem to have a positive attitude towards organ donation, they also put forward various barriers to registering as organ donor that are primarily related to fears and unease about the donation process. Distinguishing for this viewpoint were the fear that donated organs might be sold on the black market (st.17, +2), as well as unease with the possible visibility of removed organs for family members (st.12, +2) and their naked body being exposed to others during the process of transplantation (st.18, +1). Also distinguishing for this viewpoint is the disagreement with the statements saying that if you want to receive a donor organ might you need it, you should be willing to register as donor yourself (st.33, -2) and that priority should be given to registered donors in case they need an organ (st.31, -3). They also disagreed that those who have not registered as donor should not be eligible for a donor organ themselves (st.34, -3). They explained these conditions are too strong; there could be many valid reasons not to become an organ donor, such as the opinion of the family or religious reasons, and severe suffering or even death would be too serious a consequence.

People with this viewpoint doubt about the decision to register as organ donor. It is not a subject they avoid (st.23, -4), but also not one they talk about with friends (st.20, -4) or family (st.42, -1). They are generally not very well informed about the organ donation process (st.26, -2) or about how to register as organ donor (st.27, -1), and not quite so sure what information is available (st.25, +1). They think registering would not require much effort (st.22, -3) and intend to do it once they’ve made up their mind (st.28, +3), but at the same time feel that their family should eventually have the final say (st.30, +4).

Based on this description, we labelled viewpoint four “*It is a good deed, but I’m doubtful*”. A total of 5 respondents were statistically significantly associated with this factor.

3.4 Discussion

This aim of this study was to explore the prevailing viewpoints on deceased organ donation among the public in the Netherlands, and their implications for the decision to register as an organ donor. Attitudes about organ donation have been studied extensively before using survey research, but this study contributes by exploring the diversity in relations between the broad range of underlying beliefs and other constituents of people’s views on organ donation. Using Q-methodology, we found four prevailing viewpoints: “not donating your organs is a waste”, “it does not go with my religion”, “my family should decide” and “it’s a good deed, but I’m doubtful”.

The first two viewpoints are most distinct from each other, representing positions in favor and against registering as an organ donor. For people with the first viewpoint (“not donating your organs is a waste”) the potential benefits outweigh the potential barriers. The perception of self-efficacy was also relatively high in this group, with respondents indicating that they had sufficient information and knowledge about organ donation and claiming that “*as soon as I knew I wanted to become an organ donor, I signed up*”. Explanations regarding the placement of statements on the ranking grid collected during the interview revealed that for this group conversations with friends and family and encouragement by the family were important cues to action, and to a lesser extent media attention for organ donation. Answers to follow-up questions revealed that the majority of respondents who defined this viewpoint already had registered as organ donor, and otherwise stated their intention to do so in the near future.

In viewpoint two (“it does not go with my religion”), on the other hand, potential barriers clearly outweigh potential benefits, self-efficacy is low and there are few cues to action. Respondents who hold this perspective generally had little knowledge and information about organ donation. Arguably, they do not experience the need to look for such information, considering the fact they had already “*made up their mind*”. However, the qualitative materials of the interviews show that these respondents were generally dissatisfied with the type of information one can find in the media. Commercials and other information are considered superficial and understood as promotion material for registering as organ donor rather than an objective overview of different arguments and viewpoints on organ donation. As one respondent stated: “*The information I see only tells you to donate your organs because it can help other people. But it doesn’t really tell you anything about the up- and downsides of organ donation. I think the government should show examples that really resonate with all sorts of citizens. They should talk about the procedure of organ donation, but also discuss various spiritual and religious points of view*”.

Viewpoints three and four showed no clear overweight of benefits or barriers. In viewpoint three (“my family should decide”), the perceived barriers were centered on the potential harm for family members rather than for themselves. These respondents often had not registered their preference for organ donation. Not because they are indifferent, too lazy or forgot, but because they do not have a clear preference on this complicated matter. Those who had registered or stated they

intended to do so in the near future, still wanted their family to have the last word: *“I would like to become an organ donor, under the condition that my family does not object”*.

Respondents with viewpoint four (“it’s a good deed, but I’m doubtful”) were generally positive towards organ donation, but were not well informed and expressed several fears regarding the donation process. For instance, a fear of their organs ending up on the black market and an uneasy feeling about others seeing their naked body when their organs would be removed. Like those with viewpoint three, they see an important role for the family; not because of the potential trauma for family members, as in viewpoint three, but because they are uninformed and undecided. As one respondent defining this view stated: *“I should do it because it helps other people, but it does make me feel a bit uneasy. I don’t really have a preference because of these pros and cons, so my family should decide. That is less difficult for them than for me, I think, although it still may be aggravating”*.

A striking finding of this study is that people with all views appear to appreciate the benefits of organ donation, and therefore have a positive attitude towards the topic. This is quite obvious for people with the “not donating your organs is a waste” view. Those with the other three views also all consider organ donation as a good deed, but those with the “it does not go with my religion” view will not register because of their philosophy of life, those with the “my family should decide” view hand the decision to their loved ones, whereas those with the “it’s a good deed, but I’m doubtful” view appear not to be sufficiently involved with the topic to form a preference and come to a decision.

Another striking finding was the consensus among participants in the study about the neutral ranking of the statement: “information in the media helps me to determine my preference whether to become an organ donor”. Respondents explained that current information campaigns *“didn’t really help, but maybe it starts a conversation every now and then”*. Some argued that information in the media only stresses the relief in suffering for others, but does not provide any insight in the transplantation process, the communication with the family, and potential disadvantages of organ donation. Many respondents suggested it would be helpful to have a decision aid, something resembling the quizzes one can take during election times to learn which political party best represents your principles and interests. It should be noted that the interviews were conducted at the time of national elections, probably prompting respondents to make this analogy. In this context, many participants saw the ranking grid from the interview a useful tool to structure one’s thoughts about the topic of organ donation. Several respondents took a picture of their ranking grid, to further discuss the topic with their family and to deliberate themselves on the decision whether to register as organ donor.

Based on the present study one could argue that direct questions asking whether people are in favor or against organ donation may have limited information value, especially in relation to understanding decisions to register as organ donor or developing policies to influence them. Attitudes on topics like organ donation are rich and complex that need to be understood before they can be appropriately measured or influenced. As discussed, most respondents appear to have a basic positive attitude to organ donation, but in three of the four viewpoints people holding them experience different barriers that prevent them to translate this positive attitude into action.

Some limitations of this study need to be mentioned. First, we used two theoretical models and adjacent literature for identifying the variety of issues relevant for inclusion in the statement set. Although such a structured approach to the development of the statement set should make it easier to capture the full spectrum of issues, we experienced some difficulties with categorizing statements unambiguously according to these issues. In addition, although complemented with additional issues from the literature, such as ‘moral beliefs’, it may be that certain issues were missed.

We have paid extra attention to potentially missing issues relevant to respondents during the pilot study and the interviews by asking for this specifically, but no missing issues were identified by these respondents. Secondly, during the interview some respondents stated that, when in doubt about the placement of statements according to agreement, they gave higher rankings to statements that they considered more important in the decision to register as organ donor. Hence, in some cases importance was evaluated in addition to agreement as ranking criterion. Although this was not in accordance with the instructions, it is conceivable that respondents would want to give a higher rank to a statement they find important as compared to one they agree with equally but find less important. Because it concerns the relative ranking of very few statements, perhaps one column to the left or the right on the sorting grid, by a small proportion of the participants, we believe this has not affected our findings. Finally, this study identified four prevailing attitudes to organ donation but does not give information about the distribution of these views within the Netherlands. This may be relevant information if one wishes to choose optimal policies to improve donor registration. Identifying the distribution was not the aim of this study and Q-methodology is not the most appropriate method for that purpose, in particular also considering the purposive sampling of respondents. Survey research that matches respondents to these views in a representative sample of the population would be better fit for that purpose, and also verify whether any important viewpoints have been missed (Van Exel et al. 2015, Baker et al. 2010).

Our findings have some important implications for public policies aimed at promoting deceased organ donation. First, the results indicate that respondents who hold perspective three and four are in a modifiable phase. This could be understood as an opportunity to address existing fears and educate on the organ donation process. Respondents of all views indicated that current media campaigns played little to no role in their decision to register as organ donor because they are focused on the benefits for recipients and do not provide information about aspects of the process relevant to the donor, which suggests an opportunity to develop more informative media campaigns, conveying information for the various viewpoint holders.

Finally, early 2018 a law was passed in the Netherlands that will change donor registration from an opt-in system into an opt-out system effective 2020. All viewpoints show a positive attitude concerning the effects of the opt-out system, indicating that people often forget to register and that the new law will prompt people who do not want to donate their organs to make an active decision to opt out. However, viewpoint two “it does not go with my religion” expresses grave concerns regarding opt-out systems, mainly regarding family members who are not able to access or adequately understand the available information and will stay registered, thereby being considered for organ donation against their religious beliefs. People with the view “my family should decide” will be happy to know that the law provides a voice for the family, although it is not clear how the family will be involved in practice. Finally, it is not immediately clear whether

those with the “it’s a good deed, but I’m doubtful” viewpoint will appreciate the change in default. It may bring forward the fears and unease about the donation process but may also bring relief because the weight of the decision is taken away from them. Given the variety in reactions to this new law that can be expected on the basis of our findings, it seems appropriate to use the two remaining years ahead of the implementation of the law to devise information policies that take away the main concerns of people with reservations towards an opt-out system or organ donation in general, and to promote that people with different views are enabled to enact their preferences under this new system.

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>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Akhtar-Danesh, N., Brown, B., Rideout, E., Brown, M., & Gaspar, L. (2007). Use of Q-methodology to identify nursing faculty viewpoints of a collaborative BScN program experience. *Nursing Leadership*, 20(3), 67-85. <https://doi.org/10.12927/cjnl.2007.19290>
- Baker, R., van Exel, J., Mason, H., & Stricklin, M. (2010). Connecting Q & surveys: Three methods to explore factor membership in large samples. *Operant Subjectivity*, 34(1), 38-58.
- Baker, R., Thompson, C., & Mannion, R. (2006). Q methodology in health economics. *Journal of Health Services Research & Policy*, 11(1), 38-45. <https://doi.org/10.1258/135581906775094217>
- Bernhardt, A. M., & Reichensperner, H. (2014). High-risk donors: Extending our criteria in times of organ shortage. *Current Opinion in Organ Transplantation*, 19(5), 494-999. doi: 10.1097/MOT.0000000000000118
- Brown, S. R. (1980). *Political subjectivity: Applications of Q methodology in political science*. New Haven: Yale University Press.
- Brown, S. R. (1995). Q methodology as the foundation for a science of subjectivity. *Operant Subjectivity*, 18(1/2), 1-6. <https://doi.org/10.15133/j.os.1994.008>
- Bryant, L. D., Burkinshaw, P., House, A. O., West, R. M., & Ward, V. (2017). Good practice or positive action? Using Q methodology to identify competing views on improving gender equality in academic medicine. *BMJ Open*, 7(8), e015973. doi: 10.1136/bmjopen-2017-015973
- Cross, R. M. (2004). Exploring attitudes: the case for Q methodology. *Health Education Research*, 20(2), 206-213. <https://doi.org/10.1093/her/cyg121>
- Davison, S. N., & Jhangri, G. S. (2014). Knowledge and attitudes of Canadian First Nations people toward organ donation and transplantation: A quantitative and qualitative analysis. *American Journal of Kidney Diseases*, 64(5), 781-789. <https://doi.org/10.1053/j.ajkd.2014.06.029>
- Dinner, I., Johnson, E. J., Goldstein, D. G., & Liu, K. (2011). Partitioning default effects: Why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17(4), 332-341. <https://psycnet.apa.org/doi/10.1037/a0024354>

Falomir-Pichastor, J. M., Berent, J. A., & Pereira, A. (2013). Social psychological factors of post-mortem organ donation: A theoretical review of determinants and promotion strategies. *Health Psychology Review*, 7(2), 202-247. <https://doi.org/10.1080/17437199.2011.570516>

Feeley, T. H., & Moon, S.I. (2009). A meta-analytic review of communication campaigns to promote organ donation. *Communication Reports*, 22(2), 63-73. <https://doi.org/10.1080/08934210903258852>

Feeley, T. H. (2007). College students' knowledge, attitudes, and behaviors regarding organ donation: An integrated review of the literature 1. *Journal of Applied Social Psychology*, 37(2), 243-271. <https://doi.org/10.1111/j.0021-9029.2007.00159.x>

Güden, E., Çetinkaya, F., & Naçar, M. (2013). Attitudes and behaviors regarding organ donation: A study on officials of religion in Turkey. *Journal of Religion and Health*, 52(2), 439-449. <https://doi.org/10.1007/s10943-011-9490-8>

Hübner, G., Mohs, A., & Petersen, L. E. (2014). The role of attitude strength in predicting organ donation behaviour by implicit and explicit attitude measures. *Open Journal of Medical Psychology*, 3(5), 355-363.

Hyde, M. K., Knowles, S. R., & White, K. M. (2013). Donating blood and organs: Using an extended theory of planned behavior perspective to identify similarities and differences in individual motivations to donate. *Health Education Research*, 28(6), 1092-1104. <https://doi.org/10.1093/her/cyt078>

Jachimowicz, J. M., Duncan, S., & Weber, E. U. (2016). Default-switching: The hidden cost of defaults [Working paper]. Retrieved from <http://www.ssrn.com/abstract=2727301>

Johnson, E. J., & Goldstein, D. G. (2004). Defaults and donation decisions. *Transplantation*, 78(12), 1713-1716. doi: 10.1097/01.TP.0000149788.10382.B2

Johnson, E. J., & Goldstein, D. G. (2003). Do defaults save lives? *Science*, 302, 1338-1339. <https://doi.org/10.1126/science.1091721>

Kennedy, I., Sells, R. A., Daar, A. S., Guttman, R. D., Hoffenberg, R., Lock, M., Radcliffe-Richards, J., & Tilney, N. (1998). The case for “presumed consent” in organ donation. *The Lancet*, 351(9116), 1650-1652. [https://doi.org/10.1016/S0140-6736\(97\)08212-3](https://doi.org/10.1016/S0140-6736(97)08212-3)

Kitzinger, C. (1987). *The social construction of lesbianism*. London: Sage Publications.

Levitt, M., Farrell, A. M., Price, D., & Quigley, M. (2011). Educating the public to encourage organ donation. In A.M. Farrell, D. Price & M. Quigley (Eds.), *Organ shortage: Ethics, law and pragmatism* (pp. 52-66). New York: Cambridge University Press.

Manninen, D. L., & Evans, R. W. (1985). Public attitudes and behavior regarding organ donation. *Jama*, 253(21), 3111-3115. doi:10.1001/jama.1985.03350450083026

Marck, C. H., Jansen, N., Neate, S. L., Dwyer, B. M., van Haren, F., Weimar, W., & IJzermans, J. N. (2014). Comparing organ donation practices and rates between Australia and the Netherlands to detect best practices and areas for improvement. *Organs, Tissues & Cells*, 17, 175-183.

Marlais, M., Pankhurst, L., Hudson, A., Sharif, K., & Marks, S. D. (2017). UK National Registry Study of kidney donation after circulatory death for pediatric recipients. *Transplantation*, 101(6), 1177-1181. doi: 10.1097/TP.0000000000001264

McKeown, B., & Thomas, D.B. (2013). *Q methodology* (2nd ed.). Los Angeles: Sage Publications.

Mohamed, E., & Guella, A. (2013). Public awareness survey about organ donation and transplantation. *Transplantation Proceedings*, 45(10), 3469-3471. <https://doi.org/10.1016/j.transproceed.2013.08.095>

Morgan, M., Kenten, C., Deedat, S., & Donate Programme Team. (2013). Attitudes to deceased organ donation and registration as a donor among minority ethnic groups in North America and the UK: A synthesis of quantitative and qualitative research. *Ethnicity & Health*. 2013, 18(4), 367-390. <https://doi.org/10.1080/13557858.2012.752073>

Morgan, S., & Miller, J. (2002). Communicating about gifts of life: The effect of knowledge, attitudes, and altruism on behavior and behavioral intentions regarding organ donation. *Journal of Applied Communication Research*, 30(2), 163-178. <https://doi.org/10.1080/00909880216580>

Morgan, S. E. (2009). The intersection of conversation, cognitions, and campaigns: The social representation of organ donation. *Communication Theory*, 19(1), 29-48. <https://doi.org/10.1111/j.1468-2885.2008.01331.x>

Murray, L., Miller, A., Dayoub, C., Wakefield, C., & Homewood, J. (2013). Communication and consent: discussion and organ donation decisions for self and family. *Transplantation Proceedings*, 45(1), 10-12. <https://doi.org/10.1016/j.transproceed.2012.10.021>

Orentlicher, D. (2008). Presumed consent to organ donation: its rise and fall in the United States. *Rutgers Law Review*, 61(2), 295-332.

Patty, N. J., van Dijk, H. M., Wallenburg, I., Bal, R., Helmerhorst, T. J., & van Exel, J., Cramm, J. M. (2017). To vaccinate or not to vaccinate? Perspectives on HPV vaccination among girls, boys, and parents in the Netherlands: A Q-methodological study. *BMC Public Health*, 17(1), 872. <https://doi.org/10.1186/s12889-017-4879-2>

Quick, B. L., Anker, A. E., Feeley, T. H., & Morgan, S. E. (2016). An examination of three theoretical models to explain the organ donation attitude–registration discrepancy among mature adults. *Health Communication*, 31(3), 265-274. <https://doi.org/10.1080/10410236.2014.947468>

- 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. <https://psycnet.apa.org/doi/10.1037/0278-6133.16.2.183>
- Rady, M. Y., & Verheijde, J. L. (2014). The moral code in Islam and organ donation in Western countries: Reinterpreting religious scriptures to meet utilitarian medical objectives. *Philosophy, Ethics, and Humanities in Medicine, 9*(1), 11. <https://doi.org/10.1186/1747-5341-9-11>
- Risdon, A., Eccleston, C., Crombez, G., & McCracken, L. (2003). How can we learn to live with pain? A Q-methodological analysis of the diverse understandings of acceptance of chronic pain. *Social Science & Medicine, 56*(2), 375-386. [https://doi.org/10.1016/S0277-9536\(02\)00043-6](https://doi.org/10.1016/S0277-9536(02)00043-6)
- Rodrigue, J. R., Cornell, D. L., Howard, R. J. (2006). Attitudes toward financial incentives, donor authorization, and presumed consent among next-of-kin who consented vs. refused organ donation. *Transplantation, 81*(9), 1249. <https://doi.org/10.1097%2F01.tp.0000203165.49905.4a>
- Rosenstock, I. M. (1974). Historical origins of the health belief model. *Health Education Monographs, 2*(4), 328-335. <https://doi.org/10.1177/109019817400200403>
- Schmolck, P., & Atkinson, J. (2002). PQ Method Software and Manual Version 2.11.
- Shemmings, D. (2006). ‘Quantifying’ qualitative data: An illustrative example of the use of Q methodology in psychosocial research. *Qualitative Research in Psychology, 3*(2):147-165. <https://doi.org/10.1191/1478088706qp060oa>
- 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>
- Shepherd, L., & O’Carroll, R. E. (2014). Do affective attitudes predict organ donor registration? A prospective study. *Journal of Health Psychology, 19*(10), 1329-1333. <https://doi.org/10.1177/1359105313488984>
- Siegel, J. T., Navarro, M. A., Tan, C. N., & Hyde, M. K. (2014). Attitude–behavior consistency, the principle of compatibility, and organ donation: A classic innovation. *Health Psychology, 33*(9), 1084. <https://psycnet.apa.org/doi/10.1037/hea0000062>
- Smith, N. C., Goldstein, D. G., & Johnson, E. J. (2013). Choice without awareness: Ethical and policy implications of defaults. *Journal of Public Policy & Marketing, 32*(2), 159-172. <https://doi.org/10.1509/jppm.10.114>
- Stevens, C. (2013). Applying a single category implicit association test methodology to organ donation attitudes—assessing the gap between implicit and explicit beliefs [Undergraduate project, Oxford Brookes University]. <https://e-space.mmu.ac.uk/576642/>

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

Tielen, M., van Staa, A. L., Jedeloo, S., van Exel, N. J., & Weimar, W. (2008). Q-methodology to identify young adult renal transplant recipients at risk for nonadherence. *Transplantation*, *85*(5), 700-706. doi: 10.1097/TP.0b013e318166163e

Van Exel, J., Baker, R., Mason, H., Donaldson, C., Brouwer, W., & Team, E. (2015). Public views on principles for health care priority setting: Findings of a European cross-country study using Q methodology. *Social Science & Medicine*, *126*, 128-137. <https://doi.org/10.1016/j.socscimed.2014.12.023>

Van Exel, N. J., De Graaf, G., & Brouwer, W. B. (2006). "Everyone dies, so you might as well have fun!" Attitudes of Dutch youths about their health lifestyle. *Social Science & Medicine*, *63*(10), 2628-2639. <https://doi.org/10.1016/j.socscimed.2006.06.028>

Vermaire, J. H., Hoogstraten, J., van Loveren, C., Poorterman, J. H., & van Exel, N. J. (2010). Attitudes towards oral health among parents of 6-year-old children at risk of developing caries. *Community Dentistry and Oral Epidemiology*, *38*(6), 507-520. <https://doi.org/10.1111/j.1600-0528.2010.00558.x>

Wall, S. P., Plunkett, C., & Caplan A. (2015). A potential solution to the shortage of solid organs for transplantation. *Jama*, *313*(23), 2321-2322. doi:10.1001/jama.2015.5328

Watts, S., & Stenner, P. (2012). *Doing Q methodological research: Theory, method & interpretation*. London: Sage Publications. <https://dx.doi.org/10.4135/9781446251911>

Zimmerman, D., Albert, S., Llewellyn-Thomas, H., & Hawker, G. A. (2006). The influence of socio-demographic factors, treatment perceptions and attitudes to living donation on willingness to consider living kidney donor among kidney transplant candidates. *Nephrology Dialysis Transplantation*, *21*(9), 2569-2576. <https://doi.org/10.1093/ndt/gfl218>

Part 2.

EVIDENCE-BASED BEHAVIORAL POLICY

Chapter 4.

Coherence between Theory and Policy in Nudge and Boost: Is it relevant for Evidence-Based Policymaking?

D.C. Truijens

Rationality and Society, 34(3), 368-394.

4. Coherence between theory and policy in nudge and boost

Abstract

Behavioral policies Nudge and Boost are often advocated as good candidates for evidence-based policy. Nudges present or ‘frame’ options in a way that trigger people’s decision-making flaws and steer into the direction of better choices. Nudge aims to do this without changing the options themselves. Boosts also present choices in alternative ways without changing options. However, rather than steering, Boosts are aimed to increase people’s competences. Nudge and Boost originated in extensive research programs: the ‘heuristics-and-biases program’ and the ‘fast-and-frugal heuristics program’ respectively. How exactly do Nudge and Boost policies relate to the theories they originated from in the first place? Grüne-Yanoff and Hertwig labeled this a question of ‘theory-policy coherence’ and propose to use it for determining the evidence-base of Nudge and Boost. I explore the question: ‘In how far is theory-policy coherence in Nudge and Boost relevant for evidence-based policymaking?’. I argue that the implications of (weaker or stronger) theory-policy coherence are relevant in two ways. First, Grüne-Yanoff and Hertwig show that theory-policy coherence between Nudge and Boost and the research programs is not as strong as often assumed. It is crucial for the evidence-based policymaker to realize this. Assuming theory-policy coherence while it does not exist or is weaker than assumed can lead to an incorrect assessment of evidence. Ultimately it can even lead to adoption of policies on false grounds. Second, the concept of theory-policy coherence may assist the policymaker in the search and evaluation of (mechanistic) evidence. However, in order to do so, it is important to consider the limitations of theory-policy coherence. It can neither be employed as the (sole) criterion with which to determine how well grounded a policy is in theory, nor be the (sole) basis for making comparative evaluations between policies.

4.1 Introduction

Evidence-based policy prescribes that policymakers adopt those policies that are backed up with the ‘best evidence’ (e.g., Davies, Nutley and Smith 2000, Sanderson 2002, Head 2010). Although there is not one widespread definition of evidence-based policy, it generally holds that policies are based on extensive or high-quality experimental research such as randomized control trials, and meta-analyses (e.g., Cartwright, Goldfinch and Howick 2009, Cartwright and Hardie 2012, Khosrowi 2019). The behavioral policies of Nudge (Thaler and Sunstein 2008) and Boost (Grüne-Yanoff and Hertwig 2016, Grüne-Yanoff and Hertwig 2017) are often promoted as good candidates for evidence-based policy (e.g., Vlaev et al. 2016, Hertwig 2017, Schubert 2017).

A Nudge steers the chooser's behavior away from the behavior implied by the cognitive shortcoming towards the choice she would have made would she have been fully rational (e.g., healthier food choices) without changing the set of available options.

Nudges do this by invoking people's cognitive flaws, by changing features of the choice that people would generally say not to care about (e.g., position in a list, default, framing). Yet, the chooser should be able to easily choose otherwise.

See also Rebonato (2012), Grüne-Yanoff and Hertwig (2016)

Boosts can roughly be defined as having the goal of expanding people's competences to reach their own objectives without presupposing or having assumptions about what these goals are. Three classes of Boosts can be distinguished: "Policies that (1) change the environment in which decisions are made, (2) extend the repertoire of decision-making strategies, skills, and knowledge, or (3) do both."

See, Grüne-Yanoff and Hertwig (2016, 2017)

Examples of Nudges are opt-out systems for organ donation (e.g., Johnson and Goldstein 2003, Thaler and Sunstein 2008), saving programs that ask employees to save from a future salary increase rather than their current salary (Thaler and Benartzi 2004, Thaler and Sunstein 2008), and 'framing' by means of the wording of choice options in positive or negative terms (e.g., Kahneman 2003, Rothman et al. 2006, Thaler and Sunstein 2008).

Examples of Boosts are fast and frugal decision trees that replace lengthy diagnostic questionnaires for healthcare workers (e.g., Jenny et al. 2013), statistical training (e.g., Gigerenzer and Hoffrage 1995, Sedlmeier and Gigerenzer 2001) and changing the representation of statistical information from relative to natural frequencies (Gigerenzer et al. 2007, Kurz-Milcke et al. 2008).

Both Nudge and Boost policies originate from extensive research programs. Within these research programs, experimental evidence generally shows that contextual features heavily influence the choice behavior of individuals and (cognitive) theories aim to explain why they do so (e.g., Thaler et al. 1975, Kühberger and Tanner 2010). However, the research programs tell different 'stories' about people's competences and how to correct decision-making mistakes (e.g., Katsikopolous 2014). Nudge originates from the heuristics-and-biases research program (e.g., Tversky and Kahneman 1974, Thaler & Sunstein 2008) which holds that the limits on people's time, knowledge and cognitive powers prevent them from making good decisions. Boost originates from the fast-and-frugal heuristics research program (e.g., Gigerenzer and Hoffrage 1995, Gigerenzer and Todd 1999). The fast-and-frugal heuristics research program holds that despite bounds on people's time, knowledge and cognitive powers, due to the use of heuristics in environments for which these are fit, people often are in fact successful decision-makers (e.g., "ecological rationality", see for instance Goldstein and Gigerenzer 2002, Todd and Gigerenzer 2007, Todd and Gigerenzer 2012, Gigerenzer 2019).

How exactly do Nudge and Boost policies relate to the research programs they originated from in the first place? Grüne-Yanoff and Hertwig (2016) labeled this question as the problem of “theory-policy coherence” and propose to use it for determining the evidence-base of Nudge and Boost. In this article, I explore the following question: In how far is the concept of theory-policy coherence relevant for evidence-based behavioral policymaking?

I argue that there are two ways in which theory-policy coherence is relevant for the evidence-based policymaker interested in Nudge and Boost. First, I issue a warning against adopting an overly simplistic view of theory-policy coherence, one in which Nudge and Boost are fully coherent with and “naturally flow out” of the research programs they originate in. Rather, the coherence between Nudge and the heuristics-and-biases research program, or Boost and the fast-and-frugal heuristics research program, is often not as strong as might seem intuitive to assume. It is crucial for the evidence-based policymaker to realize this. For instance, falsely assuming strong theory-policy coherence can lead to the consideration of evidence that is unrelated to the policy and ultimately to justification and adoption of policies on the wrong grounds. Theory-policy coherence should be used with care, or not at all. Second, I will show that policymakers can profit from analyzing theory-policy coherence, provided they do so in a sophisticated and considered manner, taking into account the complexities of the relations between the theories and policies in question. Theory-policy coherence can be of great help to the policymaker in the search and evaluation of (mechanistic) evidence, or so I will argue.

The paper is structured as follows: Section 2 analyses a natural, intuitive way to think about theory-policy coherence in Nudge and Boost, dubbed the ‘simple view’. Section 3 discusses a more sophisticated view of theory-policy coherence as developed by Grüne-Yanoff and Hertwig (2016). I analyze how they show that Nudge and Boost are not as coherent with the heuristics-and-biases and fast-and frugal heuristics research programs as generally thought. However, I also argue that Grüne-Yanoff and Hertwig’s account of theory-policy coherence has some challenges of its own. It is too restrictive: it does not assess coherence on policy assumptions that are not necessary for implementation but are crucial for comparing and selecting one policy over the other. It is also too permissive: it allows strong theory-policy coherence on vague theories about the mechanisms underlying people’s behavior. In section 4, I argue that the concept of theory-policy coherence may still help the policymaker in the search for and evaluation of evidence, provided that neither the simple nor Grüne-Yanoff and Hertwig’s view are adopted without due consideration of their limitations. Section 5 concludes.

4.2 A first look on theory-policy coherence in Nudge and Boost

A first, intuitive, way to start thinking about theory-policy coherence in Nudge and Boost is to consider how they draw on the main theories and findings of the heuristics-and biases research program and the fast-and-frugal heuristics program, respectively.

Let us start with Nudge and the heuristics-and-biases program. According to the heuristics-and-biases program people often violate the criteria for rationality held by mainstream economics, such as consistency, transitivity, and the rules of first order logic (e.g., Thaler and Sunstein 2008, Tversky and Kahneman 1974). Experiments conducted within this research program show that minor

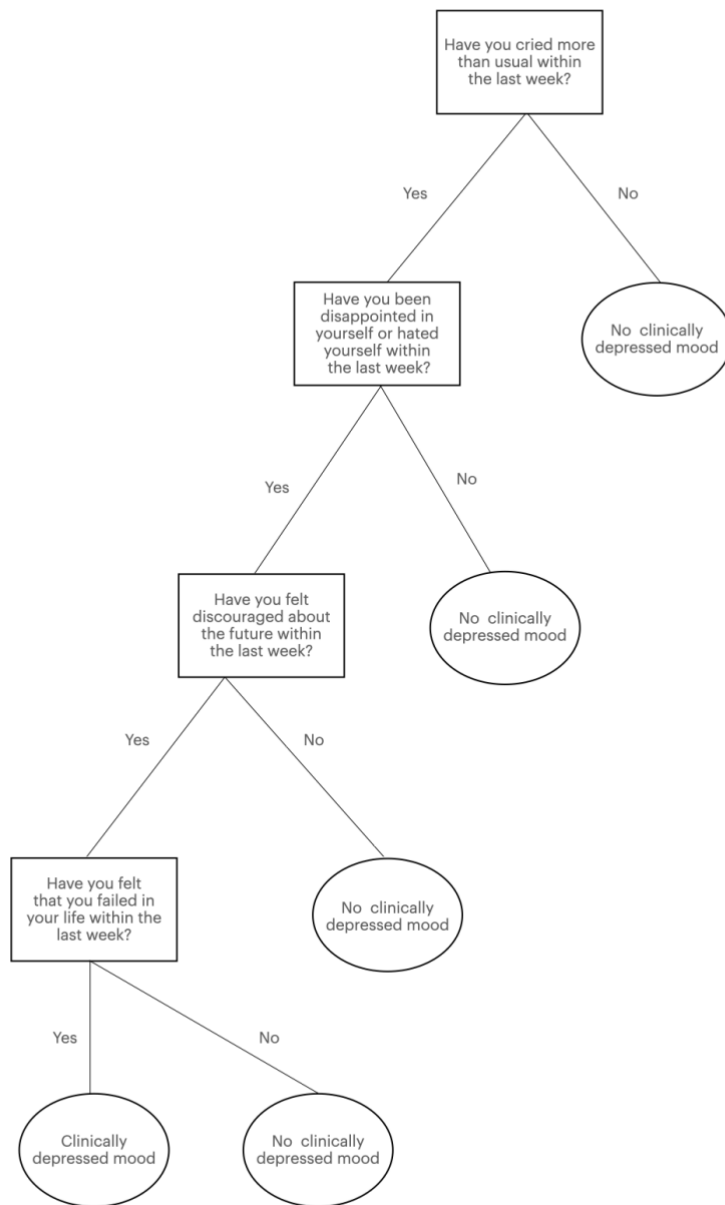
changes in the way the choice is presented influence the choices that people make. The heuristics-and-biases program also aims to explain these violations of rationality. Scholars argue that people ‘suffer’ from biases and heuristics which are ‘shortcuts’ or ‘rules of thumb’ in reasoning (e.g., Benartzi and Thaler 2007, Gilovic, Griffin and Kahneman 2002, Jolls, Sunstein and Thaler 1997). For instance, as Jolls, Sunstein and Thaler (1998, p. 1545) state: “They suffer from certain biases, such as overoptimism and self-serving conceptions of fairness; they follow heuristics, such as availability, that lead to mistakes.”

Nudge aims to help people to make better decisions as judged by themselves would they have been fully rational and is often modelled and legitimated with reference to these elements of the theoretical framework of the heuristics-and-biases program. For instance, a paradigmatic Nudge focuses on the default choice option (e.g., Johnson and Goldstein 2003, Johnson and Goldstein 2004, Thaler and Sunstein 2008, Kaiser et al. 2014). In a default Nudge there is one option that is automatically chosen if one does not choose otherwise or opt out. An example is an opt-out system for organ donation in which citizens are automatically registered as organ donors unless they actively de-register. The effect of these opt-out systems is explained by heuristics and biases such as the ‘endowment effect’ and the ‘endorsement effect’ and/or lack of willpower (e.g., Frederick 2002, Sunstein and Thaler 2003, Abadie and Gay 2006, Sunstein and Thaler 2008, Davidai et al. 2012, Dinner et al. 2011, Smith et al. 2013, Grüne-Yanoff and Hertwig 2016, 2017).

In contrast, the fast-and-frugal heuristics program does not commit to mainstream economics’ criteria for rationality (e.g., Gigerenzer and Hoffrage 1995, Gigerenzer and Todd 1999). Rather, they argue for an entirely different concept of rationality: ‘ecological rationality’. Roughly, ecological rationality concerns the fit between goals, environments, and strategies (e.g., heuristics) used. Contrary to the heuristics-and-biases program, using a particular heuristic is considered rational if using that heuristic is the right strategy in the right environment. In real life people do not have unlimited information, time and brainpower, and so using simple heuristics as rules of thumb is often efficient (Gigerenzer 1994, 1996, 2008). Moreover, because even in situations in which full information, time, and full ‘brainpower’ is available, simple heuristics sometimes still lead to better decision outcomes. That is, despite putting in less cognitive effort and not considering all information simple heuristics still obtain similar or higher accuracy, especially when applied to new information. One of the main reasons for this is ‘overfitting’ of more complex decision-making processes. This roughly entails that the decision-making process is highly accurate when applied to past data, but not to new data. This is for instance the case if too much meaning is attached to (noise from) past data when the model or decision-making process is applied to new information (e.g., Martignon and Hoffrage 2002).

A paradigmatic example of a Boost is a so-called Fast and Frugal Tree (FFT) (e.g., Gigerenzer and Todd 1999, Martignon et al. 2003, Gigerenzer 2004, Katsikoupoulous et al. 2008, Luan et al. 2011). FFT’s are small decision trees generally used to replace long and complicated decision-making processes (which take all potentially relevant information into account to perfectly ‘calculate’ the right answer). The FFT just contains a small number of yes or no questions organized in simple trees. An example of an FFT is one that helps a psychologist to determine whether a client is clinically depressed (e.g., Martignon et al. 2008, Jenny et al. 2013). For an example of an FFT, see Figure 2.

Figure 2.



Note. From “Applied decision making with fast-and-frugal heuristics”, by S. Hafenbrädl et al., 2016, *Journal of Applied Research in Memory and Cognition*, 5(2), p. 222. CC-BY-NC.

Proponents of Boost generally argue with reference to the theoretical framework of the fast-and-frugal heuristics program. They argue that if the FFT succeeds in matching the informational structure of the environment it can outperform decision-making strategies that are informationally and computationally more complex (e.g., Kozyreva and Hertwig 2019) and considers that people do not have unlimited time, information and cognitive power (e.g., Gigerenzer and Todd 1999). The idea of the FFT is that even though it does not contain a fully specified account of all nuances, it is still accurate and also more efficient than a normal questionnaire or more comprehensive decision-making process. In other words, it does not ‘steer’ people into the direction of the choice

people would have made would they have been fully rational. Rather, it aims to enhance people's decision-making competences, such that they have more accurate skills and beliefs which help them make better choices.

The above discussions may naturally lead one to adopting what I call the 'simple view': that there is a seemingly direct and strong link between theory and policy. After all, the heuristics-and-biases and fast-and-frugal research programs draw on different concepts of rationality which seem to imply policy proposals with different causal targets. That is, whereas Nudge aims to use the biases and heuristics to steer into a particular choice direction, Boost aims to eliminate the biases and apply the heuristics to help people foster and expand their competences (knowledge and skills) (Grüne-Yanoff and Hertwig 2016, 2017).

Considering the results of the heuristics-and-biases and fast-and-frugal heuristics research programs and the goals and designs of Nudge and Boost, the simple view is a natural first step to take. It shows the intuitive attraction of an analysis of Nudge and Boost that does assume the heuristics (and biases) researched in the heuristics-and-biases and fast-and-frugal heuristics programs. It furthermore is a useful benchmark for assessing the relevance of more sophisticated views on the topic, as we will discuss below.

4.3 Theory-policy coherence by Grüne-Yanoff and Hertwig

4.3.1 A more sophisticated view on theory-policy coherence

Although some authors have addressed issues concerning lack of theory-policy coherence in Nudge and Boost (e.g., Rebonato 2012, Malecka 2021), Grüne-Yanoff and Hertwig are the first to develop a comprehensive account of theory-policy coherence. Grüne-Yanoff and Hertwig (2016) demonstrate the complexity of theory-policy coherence in Nudge and Boost by explicitly analyzing whether Nudge and Boost are coherent with the theoretical research programs. They have developed an elaborate account which specifies what theory-policy coherence in Nudge and Boost should entail. In addition, they use their account of theory-policy coherence to compare theory-policy coherence in Nudge and Boost. They argue that while neither policy is fully coherent with the research program they originated from, Boost is less problematic in that sense than Nudge is. Theory-policy coherence as a concept, however, raises many questions. For instance, it raises the question of what *coherence* itself entails and between which *elements* or features of the theory we should assess for theory-policy coherence. Not all elements of the theory or research program might be equally relevant for theory-policy coherence evaluation. Furthermore, some aspects of the policies might neither be fully implied by the theoretical research programs, nor contradict them. In the following sections, first the concept that Grüne-Yanoff and Hertwig employ will be analyzed in more detail. Subsequently, aspects of theory-policy coherence which Grüne-Yanoff and Hertwig do not discuss, will be critically assessed on their relevance to the evidence-based policymaker.

4.3.2 Elements for Evaluation of Theory-Policy Coherence

Grüne-Yanoff and Hertwig base their theory-policy analysis on what they call 'necessary policy assumptions' of Nudge and Boost. Necessary policy assumptions are the implicit and explicit

assumptions underlying Nudge and Boost policies that need to be fulfilled in order for the policies to be successfully implemented. More specifically, Grüne-Yanoff and Hertwig are concerned with the question what cognitive theories Nudge and Boost assume if they aim to help people make better choices. Furthermore, they evaluate the implications of these theories for the decision-making competencies and flaws of both policymakers and the people subject to the policies.

Grüne-Yanoff and Hertwig (2016) categorize four domains on which they identify necessary policy assumptions in Nudge and Boost:

- 1) Cognitive mechanisms: error awareness and controllability,
- 2) policymakers' knowledge of people's goals and distribution in the population,
- 3) characteristics of the people the policies are trying to help: minimum competence and motivation on the part of the agents, and
- 4) characteristics of the policymaker: benevolent and less error-prone policymakers.

Grüne-Yanoff and Hertwig argue that Nudge and Boost hold contrastive necessary policy assumptions. That is, Nudge holds necessary policy assumptions that Boost typically does not hold and vice versa. For instance, according to Grüne-Yanoff and Hertwig, knowledge of people's goals and their distribution within society are necessary policy assumptions for Nudge, but not for Boost. This is because Boosts do not aim to steer people in a particular direction. They just aim to help people understand the choice situation better and expand their skills repertoire. Often, Boosts are developed on request. Grüne-Yanoff and Hertwig state that: "boost policies, unlike nudge policies, require only minimum knowledge of goals (e.g., whether statistical literacy and shared decision making are desired goals), and no knowledge of how they are distributed in the population. This is because boost policies typically leave the choice of goals to people themselves (e.g., it remains the sole decision of the statistically literate person whether or not to participate in cancer screening) or design effective tools in response to a demand expressed by professionals" (Grüne-Yanoff and Hertwig 2016, p. 166).

Nudge, in turn, steers people into the direction of the best choice option, as judged by the people themselves (Grüne-Yanoff and Hertwig 2016, Rebonato 2012, Thaler and Sunstein 2008). Therefore, policymakers should know what these options are and how they're spread within society. Furthermore, an example of a necessary policy assumption for Boost that Nudge explicitly does not have, according to Grüne-Yanoff and Hertwig, is the competence to learn new skills and the motivation to apply them. This is a necessary policy assumption for Boost since Boost requires awareness of the agents and is focused on training people to use new, better, simple heuristics. Nudge on the other hand, does not have the ambition to teach people new heuristics and therefore also does not need people to have the ability to learn new heuristics and have the motivation to apply them.

Contrary to simply assuming theory-policy coherence, Grüne-Yanoff and Hertwig tell us what the relevant elements are between which coherence is to be established and why. From the above it follows that, on their account, theory-policy coherence could exist even if some elements of the theory are not assumed by the policy, as long as these elements are not necessary policy assumptions.

4.3.3 Coherence

Grüne-Yanoff and Hertwig's evaluation of theory-policy coherence in Nudge and Boost is best illustrated by going through a few examples. A first example: Grüne-Yanoff and Hertwig argue that there is coherence with Boost's necessary policy assumptions 'controllability and awareness of the error' and the fast-and-frugal heuristics research program. For instance, in the case of an FFT, people need to actively swap their former decision-making process, such as a lengthy questionnaire, for the FFT. Hence, they argue, people need to be aware that the previously used decision-making process is faulty and that the FFT will generate better results. Furthermore, people need to be able to apply the FFT, so in that sense they need to control their faulty reasoning and make the swap. Research within the fast-and-frugal heuristics program indicates that agents often are aware of their mistakes, in the sense that they know that not every answer produced by a heuristic will be correct. In some cases, the answer produced by the heuristic can be overridden (e.g., Goldstein and Gigerenzer 2002, Pachur and Hertwig 2006, Volz et al. 2006, Grüne-Yanoff and Hertwig 2016). So here the research program holds the exact same assumption as the policy and even has empirical proof.

Furthermore, Grüne-Yanoff and Hertwig seem to argue that there is weak theory-policy coherence between Boost and the fast-and-frugal heuristics research program on the necessary policy assumptions 'motivation and minimum competence of the agent'. People who are being boosted need to actively apply the Boost. They thus need to have minimal competences to do this and be motivated to actually do it. Since the fast-and-frugal heuristics program does not say anything explicitly on these subjects, Grüne-Yanoff and Hertwig argue that regarding these assumptions theory-policy coherence is not present, but also not as profoundly absent as for some of the Nudge assumptions (Grüne-Yanoff and Hertwig 2016, p. 173-174). Hence, the policy assumption is consistent with the theoretical framework of the fast-and-frugal heuristics research program.

Lastly, and related to the previous point, Grüne-Yanoff and Hertwig argue that the necessary policy assumptions for Nudge 'knowledge of people's goals and the distribution of them within society' and 'less error prone and benevolent policymakers' are contrary to the heuristics-and-biases research program. According to them, since the heuristics-and-biases framework holds that people do not have stable preferences, it is difficult to find out what agents' goals are. Furthermore, they discuss that according to the heuristics-and-biases framework also highly educated people and other people who are supposed to be experts suffer from these decision-making flaws. Hence, based on the heuristics-and-biases research program it is highly unlikely that policymakers would not be subject to biases and heuristics (Grüne-Yanoff and Hertwig 2016 p. 172-173). Grüne-Yanoff and Hertwig show that, in this case, the policy assumption is contrary to the heuristics-and-biases framework and there is no theory-policy coherence.

Looking closely at Grüne-Yanoff and Hertwig's analysis, we can see that they use a qualified concept of *consistency*. Grüne-Yanoff and Hertwig reckon a policy coherent with the research program when the policy assumption is (one of many/multiple) implied by the theory. However, weak theory-policy coherence exists when the policy assumption is neither implied by the theory nor contrasted. There is no theory-policy coherence when the necessary policy assumption is contrary to the findings and assumptions of the theory. With this analysis, Grüne-Yanoff and

Hertwig show that the relationship of Nudge and Boost with the research programs they originated from is complex. Questions such as whether policymakers do not fall prey to the same decision-making flaws as the agents they are trying to help are crucial for evidence-based policy design. However, even though thinking in a more sophisticated way about theory-policy coherence is a crucial step forward in evaluating the theoretical commitments of Nudge and Boost, the concept is not without its pitfalls. The next section discusses crucial challenges for theory-policy coherence.

4.3.4 Challenges for Theory-Policy Coherence

4.3.4.1 Coherence as consistency

The policy assumptions underlying Nudge and Boost draw on the existence of various cognitive mechanisms underlying people's decision-making processes. Theory-policy coherence exists when necessary policy assumptions are also assumed within the theoretical frameworks of the research programs. However, the heuristics-and-biases and fast-and-frugal heuristics research programs do not always succeed in the identification of the operative mechanisms underlying the experimental results (e.g., Cubitt et al. 1998, Bohm and Lind 1993, Smith 2013, Barton and Grüne-Yanoff 2015, Grüne-Yanoff 2016). For instance, it is unclear whether people stick with defaults because of loss aversion, lack of willpower or because of the endorsement bias. This brings about at least two challenges for concepts of theory-policy coherence. First, if a policy is coherent with a theory, it remains unclear how well grounded it is in theory at that point. For instance, a policy drawing on lack of willpower can be coherent with the heuristics-and-biases program because it has identified lack of willpower as one of the possible operative mechanisms. Hence, coherence shows that, in theory, the policy could work if willpower is the operative mechanism, but it does not offer the theoretical resources to show when this is the case. In other words, there can be strong theory-policy coherence despite, or even because of, a vague theory that cannot successfully identify the operative mechanisms. After all, the more unspecific a theory is about a certain mechanism and the conditions under which it works, the easier it is for a policy to be coherent with that theory. Second, and relatedly, it is difficult to compare theory-policy coherence in policies, because it does not distinguish between policies that are coherent with a vague theory that cannot identify operative mechanisms and policies that are coherent with a more concrete theory that can do so. The question is thus how a policy could be meaningfully coherent with a theory if that theory fails in identifying the working cognitive mechanisms, such as certain biases and heuristics. Hence, theory-policy coherence is (too) permissive, because it leaves open the possibility that strong coherence would be established with a vague theory.

Furthermore, precisely because heuristics-and-biases and fast-and-frugal heuristics' theories are sometimes not sufficiently specific and cannot identify the operative mechanisms, they do not always imply *distinct* policy proposals. For example, defaults are explained differently by the fast-and-frugal heuristics framework than by the heuristics-and-biases framework. Whereas the heuristics-and-biases research program explains them by biases such as the endowment effect, the fast-and-frugal heuristics framework explains them through the endorsement effect in which the default implicitly expresses 'social information' on which option is recommended. Therefore, defaults cannot only be considered Nudges, but also be considered Boosts. This depends on the causal target of the policy as well as the underlying mechanism by which it is legitimated (see for a related point Grüne-Yanoff and Hertwig 2016). Hence, theory-policy coherence cannot make sense

of why one and the same policy is sometimes considered a Boost and other times a Nudge, without presupposing certain mechanistic theories and thus presupposing theory-policy coherence. This, in a related context, is also why Grüne-Yanoff argues for more mechanistic evidence for behavioral policies (Grüne-Yanoff 2016).

These kinds of considerations, of how well grounded a policy is in theory, go beyond an interpretation of theory-policy coherence as (a version of) consistency. Ultimately, what is needed for an operationally meaningful concept of theory policy coherence, is that policy assumptions are ‘directly’ implied by the research program. That is, for theory-policy coherence to exist, it should be established between policies and theoretical frameworks which succeed in identifying the working cognitive mechanisms, such as particular biases and heuristics. In other words, the theoretical framework should imply that the policy at issue indeed invokes a particular cognitive mechanism, and that this brings about behavior change.

4.3.4.2 Beyond necessary policy assumptions

The necessary policy assumptions that Grüne-Yanoff and Hertwig have identified are related to the debate between Nudge and Boost as competing behavioral policies. That is, the debate that centers around the question: what are the conditions under which Nudge or Boost is the more effective behavioral policy?. For instance, in this debate, Nudge is often criticized for not being able to steer people into the direction of the choice they would have made had they been fully rational. The policymakers cannot, or so critics argue, determine people’s preferences precisely because due to biases and heuristics people do not reveal their preferences through their choices (e.g., Rebonato 2012, Hertwig 2017). Boost advocates claim to evade this challenge by aiming towards a different causal target. They aim to help people have better beliefs rather than steering into a specific choice direction. They claim that in case the goals of the target population are unknown by policymakers, the goals within the population are heterogenous, or when people have conflicting goals, Boost is a better policy than Nudge. That is, when a policy aims to help people make better decisions by their own lights, Boost is the preferred policy. For instance, Hertwig (2017, p. 151) states that: “If policy-makers are uncertain about people’s goals, if there is marked heterogeneity of goals across the population or if an individual has conflicting goals, then boosting is the less error-prone intervention.”

The necessary policy assumptions that Grüne-Yanoff and Hertwig have identified are necessary for the successful execution or implementation of the policy. Without these assumptions being fulfilled, the policy goals cannot be met. Since policies are often justified based on their alleged effectiveness, assumptions like these are thus crucial for the justification of the policy. However, there are also other policy assumptions that may not be necessary for implementation but that are nevertheless crucial for justification of these policies. This is particularly prominent when comparing (behavioral) policies for selection. In order to show this, we return to the debate on whether knowledge of the target audience’s goals is necessary to improve decision-making.

Grüne-Yanoff and Hertwig have identified that ‘specific knowledge of people’s goals and the distribution of these goals within a target population’ is a necessary policy assumption of Nudge. By contrast, this is not a necessary policy assumption of Boost. The policy designer does not need to know what people’s specific goals are. This is because Boost has a different causal target, namely

to help people form more adequate beliefs and expand their decision-making skills. As we have seen, Boost advocates also claim that it is very difficult, if not impossible to know people's specific goals (e.g., Hertwig 2017, Arkes and Gassmaier 2012). Hertwig argues that "experts may even systematically misconstrue what people want for themselves...Even choosers themselves may not always be aware of their goals; sometimes they may need to work them out, and to do that, they need transparent information and the competence to process it" (Hertwig 2017, p. 152). Furthermore, as discussed, it is argued that therefore, under particular circumstances, Boost is a more effective and ethical policy than Nudge (Grüne-Yanoff and Hertwig 2016, Hertwig 2017). Hence, an assumption underlying a justification for choosing Boost over Nudge is that 'it is very difficult or even impossible to know people's specific goals and the distribution of these goals within a target population'. However, this assumption clearly is not a necessary assumption for the successful implementation of the policy. Boost could be successfully implemented even if it was easily possible to have knowledge about people's specific goals. Yet, the very assumption just quoted is a crucial policy assumption underlying the justification for Boost. Another example of an assumption underlying the justification for a policy, is the assumption underlying Nudge that people' cannot learn to overcome their biases. Even if people could learn this, Nudge could be successfully implemented. It is therefore not a necessary policy assumption. However, it is an assumption underlying an important justification for choosing Nudge (over Boost). Assumptions like these take a prominent role in the debate about whether to choose one policy over another. Assessing whether these motivations are indeed justified and there is really coherence between the policies and the research programs is important for those who want to use theory-policy coherence as a tool to compare and evaluate (behavioral) policies.

Grüne-Yanoff and Hertwig's concept of theory-policy coherence does not include other policy assumptions than those necessary for the implementation of the policy. Hence, their view of theory-policy coherence is (too) restrictive, because there can be 'complete' theory-policy coherence even though there is no coherence between 'comparative' policy assumptions underlying the justifications or motivations for the policy and the research program at issue. In other words, what is at issue here is that there also exist policy assumptions that are not necessary for the successful implementation but do have a crucial role in the debates whether to choose Nudge over Boost or vice versa.

4.4. The relevance of theory-policy coherence for evidence-based policy

4.4.1 The choice for the policymaker

Evidence-based policy prescribes that policymakers adopt those policies that have the best evidence that they work (e.g., Davies, Nutley and Smith 2000, Sanderson 2002, Head 2010). Evidence-based policy requires policies to be based on high quality empirical evidence as well as meta-analyses. Although there is not one definition or paradigm of evidence-based policy, generally it works with evidence hierarchies in which quantitative evidence is placed over qualitative evidence (e.g., Cartwright, Goldfinch and Howick 2009, Khosrowhi 2019). In particular, randomized control trials (RCT's) are considered superior over other types of evidence. As discussed, Nudge and Boost designs are often modelled after RCT's conducted within the heuristics-and-biases and fast-and-

frugal heuristics research programs. Because the results of these experiments are explained by high-level meta-analyses on the mechanisms (potentially) bringing about these results, Nudge and Boost are often considered good candidates for evidence-based policy. However, as many philosophers have argued, randomized control trials often tell you that a policy works in that specific experimental setting, but mechanistic evidence tells you why a policy works and consequently, why it would or would not work in the specific policy environment at issue (e.g., Russo and Williamson 2011). For instance, Grüne-Yanoff (2016) discusses a situation in which a default nudge has effect because of ‘weakness’ of will’ to put cognitive effort in. In this case, installing the default nudge will not work in an environment in which there already was very few cognitive effort associated with the registration. But if the mechanism behind the success of defaults is ‘loss aversion’, then the policy could work in that same environment. The mechanistic theories researched in the heuristics-and-biases and fast-and-frugal heuristics programs are thus key for the evidence-based policymaker interested in Nudge and Boost.

However, as discussed in the previous section, the relation between Nudge and Boost and the heuristics-and-biases program and fast-and-frugal heuristics program is not in any way straightforward. On some aspects, there is no theory-policy coherence. On others there is, in part. Neither of these occasions show in a meaningful way how strong the coherence is between the theoretical frameworks and the policy assumptions. Assuming theory-policy coherence while there is none, or weaker coherence than assumed leads to the motivation and adoption of policies on false grounds. For instance, if a policy is adopted based on coherence with a vague theory, then the policy might be grounded into a theory that has not been able to successfully identify the operative mechanisms and the policy cannot be legitimated based on these mechanisms. In that case, the policymaker does not have evidence that this is in fact the operative mechanism.

The challenges of theory-policy coherence leave the policymaker with a choice. The first option is this: The evidence-based policymaker can leave *any* consideration of theory(- policy coherence) aside and judge the policy on its own merits without invoking high-level mechanistic theory. She could do this by for instance conducting experiments in the specific context in which the policy is to be applied and for the specific target group of that policy. However, she will then not be able to explain why behavior occurs and extrapolate (e.g., Russo and Williamson 2011, Cartwright and Hardie 2012). The other option is that she can adopt Grüne-Yanoff and Hertwig’s view while taking its limitations into account. That is, being aware of these limitations and/or employing theory-policy coherence more carefully by looking at concrete mechanisms as well as other policy assumptions which are important in selecting a policy. However, it is crucial that the evidence-based policymaker commits to one of these options, as assuming the simple view might lead her to incorrectly present evidence for a theory unrelated to the policy.

The next section shows that if the evidence-based policymaker decides to take on theory-policy coherence and takes its limitations into account, she can use the concept constructively in the search for and evaluation of evidence. In particular, there are two main challenges related to mechanisms in the evidence-based policy discussion on Nudge and Boost that theory-policy coherence can assist with.

4.4.2 What theory-policy coherence can do for the evidence-based policymaker

4.4.2.1 Two challenges in evidence-based behavioral policy

In the literature on the evidence-base for behavioral policies, two challenges take center stage. First, as discussed, the heuristics-and-biases and fast-and-frugal heuristics research programs both draw on many alternative mechanistic explanations. However, often there is no mechanistic evidence that tells us *which* of the multiple proposed mechanisms would be at play in case Nudge and Boost are applied. This is a problem for the evidence-based policymaker. In order to know whether a policy will work in a particular environment, she needs to know which specific mechanisms are underlying the erroneous behavior the policy aims to correct. Furthermore, and in the same vein, she needs to know which mechanisms bring about the targeted behavior that the policy aims to bring about (see also sections 4.1) Furthermore, as we have seen in section 3.3, mechanistic explanations are often underlying the motivations or justifications for Nudge and Boost, and more specifically, for choosing one over the other. Nudge and Boost are both trying to help people make better decisions through invoking (biases and) heuristics (Grüne-Yanoff and Hertwig 2016, 2017). As discussed, evidence-based policy aims to select the policy with the best evidence. Therefore, mechanistic evidence is needed to assess which policy is better than the other in a particular situation in order to reach a certain goal. Moreover, the lack of mechanistic evidence raises the question if, or to what extent, the mechanisms discussed in the heuristics-and-biases and fast-and-frugal heuristics research programs are bringing about the decision-making errors at all. There might be different mechanisms underlying decision-making than the ones the heuristics-and-biases and fast-and-frugal heuristics frameworks imply, and there are different, arguably less prominent programs, referring to these different mechanisms (Rejzula et al. 2018).

Second, much of the alleged (mechanistic) evidence for Nudge's policy assumptions is contested with alleged counterevidence from the fast-and-frugal heuristics program and vice versa: evidence for policy assumptions of Boost is often contested with alleged counterevidence from the heuristics-and-biases program. For instance, many of the debates on Nudge and Boost are about questions concerning people's cognitive constitution. An example is experiments which allegedly show that people cannot learn to overcome biases or can learn to apply new, more successful, heuristics (e.g., Gilovich, Griffin and Kahneman 2002, Sunstein and Thaler 2008, Gigerenzer 2015). Relatedly, other debates are about whether people suffer from certain biases at all and whether the experiments at issue are well set-up. For instance, there is wide-spread literature on loss aversion. There is an on-going debate on whether or not the experiments conducted within the heuristics-and-biases framework indeed show that the loss aversion mechanism is indeed operative (e.g., Pachur and Hertwig 2019). For instance, amongst others, are the Asian Disease experiments (Kahneman and Tversky 1982). Gigerenzer claims that the Asian Disease example does not show that people suffer from loss aversion triggered by the framing of the example either as 'lives saved' or 'lives lost' (Gigerenzer 2015). Rather, Gigerenzer argues that in the description of the Asian Disease case one of the two alternatives is not fully specified. Due to the insufficient specification people take the wording of the example as signaling implicit information. According to Gigerenzer, it is thus not loss aversion that is the operative mechanism but rather hidden social information. This is allegedly demonstrated by experiments in which the alternatives in the Asian Disease

problem are fully specified and in which people do not suffer from loss aversion. Hence, there is, at least seemingly, conflicting evidence concerning the mechanistic theories assumed by (some) Nudges and Boosts. Relatedly, as discussed, many other debates are about questions concerning evidence for people's cognitive constitution, such as the debate on whether people can or cannot learn to overcome biases and can or cannot learn to apply new, more successful, heuristics.

4.4.2.2 Assessing (ir)relevant (mechanistic) evidence

Grüne-Yanoff and Hertwig show that Nudge and Boost are often not (fully) coherent with the heuristics-and-biases and fast-and frugal heuristics research programs respectively. As a consequence, not all evidence from the heuristics-and-biases and fast-and-frugal heuristics research programs is actually relevant for Nudge and Boost. For instance, mechanistic evidence that people can learn and apply new heuristics is not evidence for representational boosts that merely change relative into absolute frequencies. An assessment of theory-policy coherence shows why: the ability to learn and apply new heuristics is not a (necessary) policy assumption of this type of Boost. Hence, the concept of theory-policy coherence can show whether the alleged evidence is indeed evidence for the policy at issue. It therefore can have a critical, de-bunking, role: it can help show that there is no or weak theory-policy coherence and alleged evidence for theory is unrelated to the policy at issue.

4.4.2.3 Guide to what kind of mechanistic evidence is needed

If theory-policy coherence can help the evidence-based policymaker to determine in how far the presented evidence is relevant, then it can also guide the policymaker in her *search* for evidence. If there is theory-policy coherence, then evidence for a certain theory might also (but does not have to) be evidence for the policy. For instance, fast-and frugal heuristics research shows us that people often rationally select heuristics and evaluate the results they produce, and that people can furthermore learn new heuristics. Then, theory-policy coherence with Boost would tell the policymaker that evidence for that theory would be evidence for Boost. That is, as long as Boost assumes that people can learn and apply new heuristics. The policymaker could look whether such evidence is available or not. Hence, theory-policy coherence can guide the evidence-based policymaker by showing her for which theories evidence could be found. However, she needs to pay attention in the situation that the theory contains multiple mechanistic explanations but is unable to identify or explain which ones are operative under which circumstances. In those situations, the 'evidence' insufficient to consider Nudge or Boost as evidence-based.

4.4.2.4 Assessing seemingly conflicting evidence

Theory-policy coherence sheds light on to what extent seemingly conflicting evidence is indeed evidence for Nudge and Boost. Can theory-policy coherence also show to what extent the evidence is actually conflicting? Evidence from experiments that (allegedly) show that expert statisticians also use heuristics such as 'anchoring' is, if empirically correct, evidence for the heuristics-and-biases framework that 'experts also suffer from biases and use heuristics if they think intuitively'. . No matter how hard you have trained, you keep making the same mistakes if you have to think intuitively in daily life (e.g., Thaler 1985). This is evidence for Nudge's policy assumption that teaching people to reason correctly often does not make sense (e.g., Thaler and Sunstein 2008). The fast-and-frugal heuristics program's experimental evidence, on the other hand, shows that

people choose heuristics, evaluate the outcome, and choose another heuristic if they recognize the heuristic has produced an error (e.g., Pachur and Hertwig 2006, Volz 2006). Furthermore, for instance, Sedlmeier and Gigerenzer (2001) show that people can be trained to learn new decision rules. This is coherent with Boost's policy assumption that people can learn new heuristics and in principle apply these as well. Similar results are often presented as counterevidence for Nudge's policy assumption that people cannot overcome biases and the incorrect use of heuristics (e.g., Kuhberger 1995, Kuhberger and Tanner 2010, Gigerenzer 2015). However, theory-policy coherence shows us that the heuristics-and-biases research program's evidence for Nudge that experts often resort to heuristics and do not reason according to the standard picture of rationality is not conflicting with evidence from the fast-and-frugal heuristics research program for Boost that people can learn new heuristics. In fact, the heuristics-and-biases research program does not say anything about learning new heuristics (arguably, because they commit to the standard picture of rationality) and the fast-and-frugal heuristics research program does not deny, but rather endorses, that experts often make use of heuristics rather than following the rules of the standard picture of rationality. Arguably, this is because the fast-and-frugal framework entails that the application of heuristics often yields better results than reasoning according to the standard picture. Therefore, the (evidence for) the theories are not conflicting, neither fully coherent, but consistent with one another. Hence, theory-policy coherence shows what theory the evidence is for and to what extent that theory is connected to the policy assumptions. If evidence for the theories does not conflict, then that same evidence does not conflict for the policies it is coherent with.

4.4.2.5 'Crisscross coherence'

So far, we have only discussed theory-policy coherence for conventional Nudge and heuristics-and-biases program and Boost and fast-and-frugal heuristics program combinations. However, the concept of theory-policy coherence itself does not imply that it can only be established between these combinations.

What, then, about theory-policy coherence between, say, Nudge and the fast-and-frugal heuristics program and Boost and the heuristics-and-biases program? Could it make sense to consider such 'crisscross coherence'? If so, then the policymaker could, in principle, also find evidence for Boost within the heuristics-and-biases research program and for Nudge within the fast-and-frugal heuristics program. For example, take the policy assumption (in Boost) that people understand natural frequencies better than relative frequencies. That assumption is coherent with the findings of the heuristics-and-biases program that people make many relative frequencies mistakes. Although the heuristics-and-biases program might not explain the difference in people's performance between absolute and relative frequencies in terms of better or worse understanding, it does show that people's cognitive mechanisms can be faulty when dealing with relative frequencies. So, one might conclude that this important (piece of) evidence for Boost can (also) be found in experiments within the heuristics-and-biases research program – for instance, experiments that show that people give different answers to decision problems depending on whether the problem is described in absolute or relative frequencies.

Now consider a second example, this time in the opposite direction. Nudge's policy assumption that people make mistakes if they use heuristics is coherent with fast-and-frugal heuristics' theory that some heuristics are not apt for particular environments which leads to decision making errors.

Arguably, it could even be coherent with Nudge's assumption that framing a decision problem in a different way influences the choices people make. After all, changing the statistical representation of information is a form of framing. The heuristics-and-biases research program would typically explain the difference in behavior change between absolute and relative risks in terms of biases and errors. For instance, referring to a study by McDaniels (1988), Sunstein states that: "A striking study ... finds that people pervasively neglect absolute numbers, and that this neglect maps onto regulatory policy. In a similar vein, it has been shown that when emotions are involved, people neglect two numbers that should plainly be relevant: the probability of harm and the extent of harm" (Sunstein 2003, p. 766). By contrast, the fast-and-frugal heuristics program explains the behavior change in terms of better understanding (absolute frequencies) and worse understanding (relative frequencies). For instance, Gigerenzer et al. write that: "Just as absolute risks foster greater insight than relative risks do, there is a transparent representation that can achieve the same in comparison to conditional probabilities: what we call natural frequencies" (Gigerenzer et al. 2007, p. 55). Furthermore, they argue that "when the goal is to understand the size of the benefit and harm, always ask for absolute risks (not relative risks) of outcomes with and without treatment" (Gigerenzer et al. 2007, p. 60).

Crisscross coherence illustrates that while the theoretical research frameworks are, at times, very different, the policies sometimes are coherent with either research program, or the competing research program. This raises the question whether Nudge and Boost are perhaps not that different. The answer to this question is nuanced. In some instances, Nudges and Boosts have the same designs (e.g., particular defaults or statistical representations). Both experiments from the heuristics-and-biases program and the fast-and-frugal heuristics program could provide evidence for the same behavior change. In that sense, there would be crisscross coherence with experiments from both research programs. However, the heuristics-and-biases program and the fast-and-frugal heuristics program would a) provide different mechanistic theories for it and b) evaluate them differently from a normative perspective. Whether Nudge and Boost are indeed different would depend on the evidence for these theories and its coherence with the policy aims. For instance, a nudger would change the statistical representation to natural frequencies to steer people into the direction of a certain choice option. A booster would change the statistical representation to natural numbers to increase people's decision-making competences. Allegedly, he would do so without steering. Consequently, the same intervention (whether called a Nudge or a Boost) could indeed lead to behavior change. However, the fervent nudger would not claim to have increased competences, whereas an equally fervent booster would. So, if the fast-and-frugal heuristics research program too shows that people are making decisions in one direction under relative frequencies, and in another direction under absolute frequencies, then it could be coherent with Nudge's policy goals. By contrast, for the booster, coherence with the heuristics-and-biases program which holds that people are being steered is problematic. If this theory is indeed backed up with sufficient evidence, the booster's aim of increasing competences would not be met. Arguably, in that case the alleged Nudge turns out to be in fact a Boost.

One might reply to all this that the idea of crisscross coherence ignores the debates about the underlying rivaling theoretical research programs. It does. Crisscross coherence as considered here does help understand that there is important evidential overlap even between programs with different theoretical commitments. And that, in turn, can be a helpful perspective to the

policy maker who sifts through the evidence presented in the, often complex, Nudge-Boost debate and rivaling heuristics-and-biases and fast-and-frugal heuristics research programs. Thus understood, we might conclude that crisscross coherence shows that theory-policy coherence can help the policy maker understand better which evidence might be relevant and where (and in which programs), evidence can be found without pre-formed theoretical commitments. Furthermore, theory-policy coherence could help a policy maker label a proposed intervention as either a nudge or a boost (or a combination). Having said that, the prospects for crisscross coherence are naturally bound by the limitations of theory-policy coherence. In many instances it will remain the case that coherence with one theory would be in contrast, and thus mutually exclusive, with another theory.

4.4.2.6 Other theories and policies

Related to crisscross coherence, it should be noted that theory-policy coherence does not require any specific theory for Nudge and Boost to be coherent with. Therefore, it is important to note that (crisscross) theory-policy coherence can in principle also be established between Nudge and Boost and theories beyond or outside the heuristics-and-biases and fast-and-frugal heuristics research programs. Grüne-Yanoff and Hertwig also indicate this when they discuss Boost's necessary policy assumption 'sufficient motivation of the agent to apply the Boost'. Since the fast-and-frugal heuristics framework does not say anything about the extent to which agents are motivated to learn and apply new heuristics, the evidence-based policy maker could also look for (motivational) theories outside the heuristics-and-biases and fast-and-frugal heuristics research programs that show that agents are motivated to apply Boosts or other similar newly learned tools (Grüne-Yanoff and Hertwig 2016, 2017). In that case, Boost would be coherent with another theory and the evidence-based policy maker could search for evidence supporting that theory at issue in order to establish evidence for the necessary policy assumption that agents need to have minimal motivation to apply Boost.

Furthermore, and in the same vein, other (behavioral) policies than Nudge and Boost could be coherent with the heuristics-and-biases and fast-and-frugal heuristics programs. For instance, Reijula et al. (2018) and Ehrig et al. (2015) propose an alternative behavioral policy to Nudge and Boost that intervenes on the institutional rather than at the individual level: 'mechanism and norm design'. They argue that Nudges and Boost are often employed to generate social outcomes, but that individual rationality often does not translate into social welfare, as problems such as 'tragedy of the commons' illustrate. Arguably, in this context, they implicitly make a case against theory-policy coherence of Nudge and Boost's social goals with classic macro-economic problems. Mechanism and norm design, by contrast, take the social outcome as a starting point. Ehrig et al. (2015) argue that more organ donations are not always better because of, for instance, costs associated with the medical procedures. As an alternative policy, they refer to an organ donation matching tool as developed by Roth, Sönmez, and Ünver (2004) in which recipients and donors are more efficiently matched. In addition, there are also donation options that provide higher priority. Policies like these may have a stronger coherence with macro-economic theories than with the heuristics-and-biases or fast-and-frugal heuristics program. However, it is conceivable that there are combinations possible with for instance nudges to get registrations for such a matching system. In that case, theory-policy coherence could exist with both macro-economic theories as well as with the heuristics-and-biases framework. However, it should be stressed that theory-policy

coherence should not be used for cherry-picking. That said, it allows policies to be evaluated against their coherence with a variety of academic disciplines and research programs.

4.5 Conclusions

Assuming a tight relation between Nudge policies and heuristics-and biases theory and Boost policies and fast-and-frugal heuristics theory is problematic. In various instances, Nudge and Boost assume decision-making competencies and flaws that are only merely broadly consistent with, or even contrary to, the theoretical frameworks of the heuristics-and-biases and fast and-frugal heuristics research programs. Furthermore, even in cases where they are ‘fully coherent’ with the heuristics-and-biases and fast-and-frugal heuristics research programs, this does not imply that the policy is strongly based on theory. For instance, evidence for a vague theory that is coherent with the policy but unable to explain which of multiple possible alternative mechanisms is operative is not helpful to determine whether the policy will work in a given situation. Furthermore, the evidence-based policymaker should evaluate coherence with all policy assumptions that may justify the selection of one policy over another, rather than only with policy assumptions that are necessary for the success of its implementation. The challenges of theory-policy coherence leave the evidence-based policymaker with a choice. The choice implies that the evidence-based policymaker should be aware of where and when she brings in theory, and if she does, examine coherence with the policy. If the evidence-based policymaker does bring in theory, theory-policy coherence can aid in two major challenges she faces regarding Nudge and Boost: the search for (relevant) mechanistic evidence and the examination of seemingly conflicting evidence. Regarding the first, theory-policy coherence can guide the evidence-based policymaker by showing which mechanistic theories evidence should be found. In addition, it can help the policymaker assess which evidence is coherent, and thus relevant for the policy, and which not. Regarding the second, by showing which evidence is relevant and which not, theory-policy coherence can also help the policymaker to see whether seemingly conflicting evidence is in fact conflicting. Thus is the relevance of coherence between theory and policy for Nudge and Boost in evidence-based policymaking.

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>
- Arkes, H. R., & Gaissmaier, W. (2012). Psychological research and the prostate-cancer screening controversy. *Psychological Science*, 23(6), 547-553. <https://doi.org/10.1177/0956797612437428>
- 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>
- Benartzi, S., & Thaler, R. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives*, 21(3), 81-104. doi: 10.1257/jep.21.3.81
- Cartwright, N., Goldfinch, A., & Howick, J. (2009). Evidence-based policy: Where is our theory of evidence?. *Journal of Children's Services*, 4(4), 6-14. <https://doi.org/10.5042/jcs.2010.0017>
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. New York: Oxford University Press. <https://doi.org/10.5042/jcs.2010.0017>
- Davies, H. T., Nutley, S. M., & Smith, P. C. (2000). *What works?: Evidence-based policy and practice in public services*. Bristol: Policy Press.
- Davidai, S., Gilovich, T., & Ross, L. D. (2012). 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>
- Dinner, I., Johnson, E.J., Goldstein, D.G., & Liu, K. (2011). Partitioning default effects: Why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17(4), 332–341. <https://psycnet.apa.org/doi/10.1037/a0024354>
- Ehrig, T., Katsikopoulos, K. V., Kuorikoski, J., Poyhonen, S., & Sunder, S. (2015). Limitations of behaviorally informed policy under social interaction. Available at SSRN 2685912. <https://dx.doi.org/10.2139/ssrn.2685912>
- Frederick, S. (2002). Automated choice heuristics. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 548–558). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511808098.032>
- Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective Probability* (pp. 129-161). New York: Wiley.

- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592-596. <https://psycnet.apa.org/doi/10.1037/0033-295X.103.3.592>
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. <https://psycnet.apa.org/doi/10.1037/0033-295X.102.4.684>
- Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96. <https://doi.org/10.1111/j.1539-6053.2008.00033.x>
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*(1), 20-29. <https://doi.org/10.1111/j.1745-6916.2008.00058.x>
- Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology*, *6*(3), 361-383. <https://doi.org/10.1007/s13164-015-0248-1>
- Gigerenzer, G. (2021). Axiomatic rationality and ecological rationality. *Synthese*, *198*(4), 3547-3564. <https://doi.org/10.1007/s11229-019-02296-5>
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press. <https://psycnet.apa.org/doi/10.1017/CBO9780511808098>
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75. <https://psycnet.apa.org/doi/10.1037/0033-295X.109.1.75>
- Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics & Philosophy*, *32*(3), 463-483. <https://doi.org/10.1017/S0266267115000425>
- Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory?. *Minds and Machines*, *26*(1), 149-183. <https://doi.org/10.1007/s11023-015-9367-9>
- Hafenbrädl, S., Waeger, D., Marewski, J. N., & Gigerenzer, G. (2016). Applied decision making with fast-and-frugal heuristics. *Journal of Applied Research in Memory and Cognition*, *5*(2), 215-231. <https://doi.org/10.1016/j.jarmac.2016.04.011>

- Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), 973-986.
<https://doi.org/10.1177/1745691617702496>
- Head, B. (2010). Evidence-based policy: principles and requirements. In *Strengthening evidence-based policy in the Australian Federation, volume 1: Preceedings* (pp. 13-26).
- Hertwig, R. (2017). When to consider boosting: Some rules for policy-makers. *Behavioural Public Policy*, *1*(2), 143-161. <https://doi.org/10.1017/bpp.2016.14>
- Johnson, E.J., & Goldstein, D.G. (2003). Do defaults save lives? *Science*, *302*, 1338-1339.
<https://doi.org/10.1126/science.1091721>
- Jolls, C., Sunstein, C. R., & Thaler, R. H. (1998). A behavioral approach to law and economics. *Stanford Law Review*, *50*, 1471.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–210). Cambridge: Cambridge University Press.
- Kaiser, F., Arnold, O., & Otto, S. (2014). Attitudes and defaults save lives and protect the environment jointly and compensatorily: Understanding the behavioral efficacy of nudges and other structural interventions. *Behavioral Sciences*, *4*(3), 202-212.
<https://doi.org/10.3390/bs4030202>
- Katsikopoulos, K. V., Pachur, T., Machery, E., & Wallin, A. (2008). From Meehl to Fast and Frugal Heuristics (and Back) New Insights into How to Bridge the Clinical—Actuarial Divide. *Theory & Psychology*, *18*(4), 443-464. <https://doi.org/10.1177/0959354308091824>
- Katsikopoulos, K. V. (2014). Bounded rationality: the two cultures. *Journal of Economic Methodology*, *21*(4), 361-374. <https://doi.org/10.1080/1350178X.2014.965908>
- Khosrowi, D. (2019). Extrapolation of causal effects—hopes, assumptions, and the extrapolator’s circle. *Journal of Economic Methodology*, *26*(1), 45-58.
<https://doi.org/10.1080/1350178X.2018.1561078>
- Kozyreva, A., & Hertwig, R. (2021). The interpretation of uncertainty in ecological rationality. *Synthese*, *198*, 1517–1547. <https://doi.org/10.1007/s11229-019-02140-w>
- Kuhberger, A. (1995) The framing of decisions: A new look at old problems. *Organizational Behavior and Human Decision Processes*, *62*(2), 230-240. <https://doi.org/10.1006/obhd.1995.1046>
- Kuhberger, A. & Tanner, C. (2010). Risky choice framing: task versions and a comparison of prospect theory and fuzzy-trace theory. *Journal of Behavioral Decision Making*, *23*(3), 314-329.
<https://doi.org/10.1002/bdm.656>

- Jenny, M. A., Pachur, T., Williams, S. L., Becker, E., & Margraf, J. (2013). Simple rules for detecting depression. *Journal of Applied Research in Memory and Cognition*, 2(3), 149-157. <https://doi.org/10.1016/j.jarmac.2013.06.001>
- Johnson, E. J., & Goldstein, D. G. (2004). Defaults and donation decisions. *Transplantation*, 78(12), 1713-1716. doi: 10.1097/01.TP.0000149788.10382.B2
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, 58(9), 697. <https://psycnet.apa.org/doi/10.1037/0003-066X.58.9.697>
- Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication: Graphical and analog tools. *Annals of the New York Academy of Sciences*, 1128(1), 18-28. <https://doi.org/10.1196/annals.1399.004>
- Luan, S., Schooler, L. J., & Gigerenzer, G. (2011). A signal-detection analysis of fast-and-frugal trees. *Psychological Review*, 118(2), 316. <https://psycnet.apa.org/doi/10.1037/a0022684>
- Malecka, M. (2021). Knowledge, behaviour, and policy: questioning the epistemic presuppositions of applying behavioural science in public policymaking. *Synthese*, 199(1), 5311-5338. <https://doi.org/10.1007/s11229-021-03026-6>
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52(1), 29-71. <https://doi.org/10.1023/A:1015516217425>
- Martignon, L., Vitouch, O., Takezawa, M., & Forster, M. R. (2003). Naive and yet enlightened: From natural frequencies to fast and frugal decision trees. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspective on reasoning, judgment, and decision making* (pp. 189-211). Chichester: Wiley. <https://doi.org/10.1002/047001332X.fmatter>
- Martignon, L., Katsikopoulos, K. V., & Woike, J. K. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology*, 52(6), 352-361. <https://doi.org/10.1016/j.jmp.2008.04.003>
- McDaniels, T. L. (1988). Comparing expressed and revealed preferences for risk reduction: Different hazards and question frames. *Risk Analysis*, 8(4), 593-604. <https://doi.org/10.1111/j.1539-6924.1988.tb01203.x>
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 983-1002. <https://psycnet.apa.org/doi/10.1037/0278-7393.32.5.983>
- Rebonato, R. (2012). *Taking liberties: A critical examination of libertarian paternalism*. New York: Palgrave Macmillan.

- Reijula, S, Kuorikoski, J, Ehrig, T, Katsikopolous, K, Sunder, S. (2018). Nudge, Boost or Design? Limitations of behaviorally informed policy under social interaction. *The Journal of Behavioral Economics for Policy* 2(1), 99-105. <https://dx.doi.org/10.2139/ssrn.2685912>
- Roth, A. E., Sönmez, T., & Ünver, M. U. (2004). Kidney exchange. *The Quarterly Journal of Economics*, 119(2), 457-488. <https://doi.org/10.1162/0033553041382157>
- Rothman, A. J., Bartels, R. D., Wlaschin, J., & Salovey, P. (2006). The strategic use of gain-and loss-framed messages to promote healthy behavior: How theory can inform practice. *Journal of Communication*, 56(Suppl_1), S202-S220. <https://doi.org/10.1111/j.1460-2466.2006.00290.x>
- Russo, F., & Williamson, J. (2011). Epistemic causality and evidence-based medicine. *History and Philosophy of the Life Sciences*, 33(4), 563-581.
- Sanderson, I. (2002). Evaluation, policy learning and evidence-based policymaking. *Public Administration*, 80(1), 1-22. <https://doi.org/10.1111/1467-9299.00292>
- Schubert, C. (2017). Green nudges: Do they work? Are they ethical?. *Ecological Economics*, 132, 329-342. <https://doi.org/10.1016/j.ecolecon.2016.11.009>
- Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, 130(3), 380-400. <https://psycnet.apa.org/doi/10.1037/0096-3445.130.3.380>
- Sunstein, C. R. (2003). Hazardous heuristics. *The University of Chicago Law Review*, 70(2), 751-782. <https://doi.org/10.2307/1600596>
- Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 70(4), 1159-1202. <https://doi.org/10.2307/1600573>
- Thaler, R.H. (1985) Illusions and mirages in public policy. In V.T. Covello, J.L. Mumpower, P.J.M. Stallen & V.R.R. Uppuluri (Eds.), *Environmental impact assessment, technology assessment, and risk analysis* (pp.567-581). Berlin: Springer. <https://doi.org/10.1007/978-3-642-70634-9>
- Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1), S164-S187. <https://doi.org/10.1086/380085>
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.
- Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, 16(3), 167-171. <https://doi.org/10.1111/j.1467-8721.2007.00497.x>

Todd, P. M., & Gigerenzer, G. E. (2012). *Ecological rationality: Intelligence in the world*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195315448.001.0001

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>

Vlaev, I., King, D., Dolan, P., & Darzi, A. (2016). The theory and practice of “nudging”: changing health behaviors. *Public Administration Review*, 76(4), 550-561. <https://doi.org/10.1111/puar.12564>

Volz, K. G., Schooler, L. J., Schubotz, R. I., Raab, M., Gigerenzer, G., & von Cramon, D. Y. (2006). Why you think Milan is larger than Modena: Neural correlates of the recognition heuristic. *Journal of Cognitive Neuroscience*, 18(11), 1924–1936. <https://doi.org/10.1162/jocn.2006.18.11.1924>

Chapter 6.

Conclusions

6. Conclusions

6.1 Back to where it began

This thesis started with the ambition to have a closer look at the evidence-base for behavioral policies. Behavioral policies have gone through enormous popularization in public science and policymaking. In part, this is because behavioral policies are justified based on extensive research programs in behavioral psychology and economics. More specifically, behavioral policies aim to improve agents' decision-making (outcomes), thereby drawing on both experimental behavioral research which shows that people are often influenced by the mere presentation of choice options, and normative theories about the (ir)rationality of agents' decision making.

In light of this, I investigated to what extent the justifications for behavioral policies as based on an alleged broad and robust evidence-base held up. Particularly, I've aimed to increase understanding on which basis rationality claims about agents' decision-making processes were made, and why, and how these claims based on experimental evidence played a role in the justification of behavioral policies. Ultimately, behavioral policies draw on interpretations of experimental research by analysts, whether they are policy makers or researchers. These analysts are interpreting agents' decision-making processes and views on the decision problem. The thesis therefore has a specific focus on common analyses of agents' own views of the decision problem and the choice options they are presented with in behavioral interventions. Relatedly, it aims to connect highly theoretical research in which these interpretations of agents' behavior are grounded with the more practical realm of evidence-based policymaking. That is, rather than debating the theoretical grounds for policy assumptions themselves, I have chosen to focus on showing the implications of the (often weak or controversial) evidence-base for behavioral policies.

Rather than merely criticizing or to issue a warning about these evidential challenges, I have also aimed to shed light on what behavioral policy makers can (still) do with the evidence there is. That is, I have investigated how policy makers can improve their search and evaluation of scientific evidence. I have done this through a mix of more theoretical and applied research projects, both philosophical and empirical. Analyzing behavioral policies this way helps us discuss, critically evaluate, and further develop behavioral policies in light of their theoretical assumptions as well as the practical goals they aim to achieve.

In the remainder of these conclusions, I will first recap the findings regarding with what I dub 'evidential challenges' for behavioral policies. Second, I will discuss the implications for evidence-based policy.

6.2 Evidential challenges for behavioral policies

In *chapter 2*, I started out by discussing what I have dubbed the 'Mechanistic Evidence Challenge' for behavioral policies. Various philosophers of science, such as Grüne-Yanoff, Gigerenzer, and Marchionni and Reijula have pointed out that behavioral policies draw on (implicit) assumptions about agents' cognitive mechanisms. For instance, some behavioral policies draw on the assumption that default effects are brought about by loss aversion, weakness of will, and/or other

biases or heuristics. These philosophers have argued that 1) research suggests multiple competing mechanisms in parallel, while 2) it is not clear which one of those, if any at all, are operative, 3) that which particular mechanism is being assumed matters for extrapolation from the experimental results to the policy environment, and that 4) therefore more mechanistic evidence needs to be gathered. This ‘mechanistic evidence challenge’ can be summarized as behavioral policies requiring more mechanistic evidence for the purpose of extrapolation.

In the same chapter, building on the mechanistic evidence challenge, I proposed the “Individuation Challenge”. To do so, I investigated the famous, paradigmatic case called ‘Surgery’. Surgery contains a series of behavioral experiments showing that framing or describing the same choice options in negative or positive terms brings about behavior change. Many behavioral policies have been modelled after these framing cases. However, thorough analysis of the experiments indicates that while the experiments show behavior change, they do not necessarily show that agents were (irrationally) inconsistent. Rather, to interpret the behavior of agents as irrationally inconsistent, one needs to draw on a very specific interpretation (or theory) of how options should be individuated or described. This is even so according to mainstream economics’ criteria for rationality. The interpretations draw on an analysis of assumptions about how agents see choice options and whether them doing so is rationally justified or not. Put simply, for analysts to interpret behavior change as inconsistent behavior, they need to establish that the choice options did not meaningfully change under the different frames. All this, as it turns out, is highly controversial. The interpretations of Surgery and similar behavioral experiments draw on specific assumptions and norms about which features of the choice situation *do matter* and *should matter* to agents. These assumptions draw on high-level decision theory. And these assumptions are very controversial, even within mainstream economics. Furthermore, there are also plausible alternatives offered by competing research programs such as the Simple Heuristics program. Consequently, the common interpretation of Surgery also presupposes, rather than merely establishes, evidence about agents’ cognitive processes and beliefs about choice options. This entails, but goes beyond, the mechanistic evidence challenge. The analysis of Surgery shows that it is not only questionable whether experimental results can be extrapolated. It is also questionable whether agents showed (irrationally) inconsistent behavior in experiments in the first place. This, or so I have argued, poses what I call the “Individuation Challenge” for behavioral policies. The challenge holds that for behavioral policies to be justified as evidence-based, an (explicit) theory of how options should be described for analysis should be developed. This includes the role of agents’ views on the choice options. Without such theory, behavioral policies cannot be accurately justified as helping irrationally inconsistent agents make better choices (as judged by their own lights).

One way to gather more insight in agents’ own views regarding choice options is to extend quantitative research with qualitative insights. The aim of *chapter 3* was to gain insights in how agents think about the choice whether to register as an organ donor or not. Defaults for organ donation registration are often advocated on the basis that people show positive attitudes towards donation in quantitative research. Furthermore, they are often advocated as drawing on cognitive ‘failures’ on part of the agent. Examples of these failures are status quo biases, weakness of will, endowment bias, and so on. However, it is unclear how agents really think about the individual choice to register as organ donors. Furthermore, it is unclear whether defaults would indeed trigger the aforementioned cognitive mechanisms. The findings of *chapter 3* were a result of a mixed

method (Q-methodology) study amongst Dutch citizens about their attitudes towards organ donation registration. Rather than taking individual questions and answers, Q methodology aims to connect individual statements and analyze them in relation to each other. The study showed that agents' views can generally be categorized in four different views. At least two of these views, namely 'my family can decide' and 'It's a good deed, but I'm doubtful', provide strong indications for why agents stick with defaults without being irrational. For instance, arguably, agents may stick with defaults because they let their families make the final decision in any case. Opt-in or opt-out systems do not make a meaningful difference for them. After all, family members will be able to decide what to do with the body eventually anyway. In the same vein, agents who aim to think more carefully about the decision might be registered under an opt-out system but may opt-out later.

Chapter 4 discusses a study by Grüne-Yanoff and Hertwig who show that behavioral policies draw on theoretical assumptions that may even conflict or be incoherent with the research programs they originated from. Examples of these are assumptions about analysts' being able to determine what is the best choice option for agents (behavioral preference elicitation) and agents being aware of and being able to control the bias that is allegedly triggered by the intervention. The goal of chapter 4 was to analyze what Grüne-Yanoff and Hertwig call the concept of theory-policy coherence (TPC) and evaluate to what extent it is relevant for evidence-based policymaking. I have argued that while TPC is crucial in understanding that the connection of the policies with the theories they are often justified by is often weaker than assumed, it also has important limitations. On the one hand, TPC is too permissive: a policy can be coherent with vague theories about the mechanisms underlying people's behavior. On the other hand, TPC is too restrictive: the concept of theory-policy coherence as developed by Grüne-Yanoff and Hertwig considers the (implicit) assumptions of the policies themselves (that is, those that are necessary for their implementation), but does not take into account the wider range or context of justifications of motivations that are provided for those policies. For instance, TPC does consider the fact that behavioral policies assume that their designers are not as prone to biases and heuristics as the target group. However, TPC does not consider an often-voiced justification for Nudge over other (behavioral) policies that people can hardly learn from their mistakes. This, arguably, could be understood as another challenge for behavioral policies. Namely, in order for behavioral policies to be adequately assessed on their evidence base, both the policy assumptions and wider justifications for these policies need to be assessed on their coherence with the research programs they are motivated by. Therefore, I have concluded that TPC can neither be employed as the (sole) criterion with which to determine how well grounded a policy is in theory, nor be the (sole) basis for making comparative evaluations between policies. It is crucial for the evidence-based policy maker to be aware of this, as assuming theory-policy coherence while it does not exist or is weaker than assumed, can lead to an incorrect assessment of evidence and ultimately to adoption of policies on false grounds.

The challenges discussed and identified in this thesis imply that behavioral policies are based on controversial (implicit) assumptions. While some of these assumptions are controversial because of deeply entrenched debates within the theoretical research programs they stem from, others are controversial because there are alternative plausible interpretations from competing research programs. Yet again, neither of these assumptions seem to have a strong connection with the research programs the policy interventions originated from. Rather, they sometimes turn out to be

incoherent or even inconsistent with them. This thesis thus teaches us that while behavioral policies are often seen as (good candidates for) evidence-based policies, the evidence is often not as strong and straightforward as it might seem. Indeed, the evidence is often subject to heavy interpretations. These interpretations, in turn, are based on controversial assumptions.

One, but not the only, reason for the controversial aspects of behavioral evidence is that the (interpretations of) experimental evidence is deeply rooted in normative theories of behavior. That is, when behavioral policies are advocated, rationality claims are often explicitly or implicitly made. These are all ‘value claims’ about behavior. Moreover, the competing notions of rationality, experimental methodologies, and expert disagreement on interpretations of behavioral policies are a rich ground for the broader values in science debate. The behavioral policy debate, and more particularly the evidential challenges as I have discussed and identified, are all clear cases of where and when epistemic and non-epistemic values are intertwined with scientific results. We can see that this goes far beyond often-mentioned examples such as choice of methodologies, communicating probabilities, and determining sufficiency of evidence. For instance, the individuation challenge has shown that merely identifying inconsistencies is a normative endeavor in which one needs to decide what the ‘right way’ of viewing choice options is. In this context, the evidential challenges I have discussed and proposed in this thesis shed light on several ‘value moments’ that often go unnoticed both in the theoretical literature and more practical communication and use of behavioral research.

6.3 Implications for evidence-based policy

From the evidential challenges discussed in the various chapters of this thesis, one might be inclined to conclude that behavioral policies are not evidence-based. A strict advocate of the value free ideal of science could argue that we should refrain from developing and implementing behavioral policies altogether. However, although issues concerning sufficiency of evidence, value ladenness and expert disagreement might be more prominent in behavioral policies, they are not unique in evidence-based policy. While one could still argue that science should strive towards a value free ideal, this might neither be feasible nor desirable for behavioral research and evidence-based policy generally. After all, the value of behavioral research is also in its interpretation of behavior and its connection with the question of how welfare benefits could be created or optimized. Taking this into account, the underlying ambition of this thesis is not only to ‘warn’ against behavioral policies or criticize them. Rather, in this thesis I have discussed the issues with a more constructive angle for policy makers in mind: I have also evaluated what the policy maker or practically interested researcher *can do* to engage the relevant evidence and all its underlying assumptions more adequately.

What implications do the behavioral challenges have for evidence-based policy? Chapter 4 showed that behavioral policies often draw on policy assumptions and are justified based on assumptions that are not, or only weakly, coherent with the research programs they are allegedly based upon. Although I have shown that an evaluation of theory-policy coherence cannot be the sole criterion for evidence-based policy, chapter 4 also contributes to the constructive aim of the thesis by showing that it may help the policy maker in identifying which evidence needs to be searched for, and where. More specifically, it can help to distinguish between those policy assumptions

underlying the motivations or justifications for the policy and policy assumptions necessary for the successful implementation of the policy and explicitly seek for coherence with specific, rather than vague, mechanisms. That is, to look for the mechanisms a policy assumes to be at work in the specific situation and target population where the policy is to be implemented. If the evidence-based policy maker does so, theory-policy coherence can aid in two major ways: the search for (relevant) mechanistic evidence and the examination of seemingly conflicting evidence. Regarding the first, theory-policy coherence can guide the evidence-based policy maker by showing which mechanistic theories evidence should be found. In addition, it can help the policy maker assess which evidence is coherent, and thus relevant for the policy, and which not.

Regarding the search for mechanistic evidence, chapters 2, 4 and 5 have argued that policy makers and researchers alike should aim to find more mechanistic evidence for presupposed cognitive processes. Particularly, I have argued for more research on the views people have about the choice options. For instance, in case of organ donation, it is important to realize for policy makers that positive attitudes towards organ donation do not necessarily imply an intention or willingness to register as an organ donor. Agents' reasons for registering as organ donors and/or sticking with default positions, whatever they may be, are complex and nuanced. Chapter 3 showed that there are four main target groups of agents in the Netherlands, of which at least two may stick with defaults while they may not in the end donate their organs. I have argued that more research should be done about different target groups and their characteristics. Furthermore, the results discussed in chapter 3 indicate that the current number of registrations in an opt-out system may be much higher than the final number of people actually (willing to) donate their organs. This has crucial implications for the understanding of the efficiency of the policy and the estimation of available organs. In other words, even a 'classic success case' for behavior policies might not be as successful in the long term as generally advocated. Agents might not stick with organ donation defaults for the reasons that are commonly stated. Relatedly, regarding (seemingly) conflicting evidence, the individuation challenge in particular shows that there is a high level of expert disagreement about the interpretation of behavioral experiments generally. Not only do experts disagree about which cognitive processes are operative. They also disagree about research methodologies for decision analysis and normative rationality standards. The thesis, and specifically chapter 2, show that merely gathering more mechanistic evidence does not suffice. Rather, a theory of options needs to be developed which shows how analysts, including policy makers, should describe the choice options when analyzing behavior change. Such a theory could depend on the goal of the policy (maker). Arguably, the more the policy aims to correct alleged irrationalities rather than just steer towards certain behavior, the more important the role of agents' views in the theory of options becomes.

Chapter 5 was devoted entirely to the question of what the policymaker should do in light of the evidential challenges discussed in this thesis. I argued that behavioral policy makers should be aware of various behavioral policies, alternatives and the evidential challenges and scholarly disagreements associated with them. This often is not the case. Behavioral insight teams and public debates about behavior policies are generally focused on the Heuristics and Biases research program and the behavioral policy Nudge that originated from it. Alternative research programs such as the Simple Heuristics program and behavioral policies such as Boost are not as commonly known. Debates about the policy assumptions and sufficiency of evidence should therefore benefit from interaction between policymakers and research(ers) from a variety of research traditions or schools. In the last

leg of this thesis, I have therefore presented the '3xE framework'. By using this framework, existing policy guidelines and tools can be expanded and modified such that behavioral evidence, and the implicit values associated and underlying these are made more explicit and can be debated. Examples of are different compositions of expert panels, literature reviews or evidence gap maps, and explicit evaluation of multiple (behavioral) policy alternatives.

Through this thesis, I hope to have shown that the evidence for behavioral policies is often complex, controversial, and weaker than assumed. Specifically, I have aimed to shed light on implicit policy assumptions, particularly those presupposing agents' views on the decision problems they face and the normative evaluations of those. Ultimately, I hope to have shown that often highly abstract issues debated in the philosophy of science and economic (behavioral) theory do have actual bearing in the practical realm of evidence-based policymaking. And, last but not least, that shedding light on those, often value-laden, issues is of significance to policymakers and may be mitigated through more adequate use of evidence-based behavioral policy practices and guidelines.

Summary

Recent discoveries in psychology and behavioral economics show that small, seemingly irrelevant, changes in how options are presented have large effects on the choices that people make. In this context, people often seem to be irrational decisionmakers. These so-called “behavioral insights” from behavioral economics and psychology have inspired researchers and policymakers alike to develop interventions, which aim to help people make better decisions for society and for themselves. Smaller plates lead to better sticking to diets, default opt-out registration systems lead to more people signing up as organ donors, and signs telling us about the neighbors spending less minutes in the shower leads people to save water. These changes in the choice environment are relatively simple and cost efficient. Behavioral insights even catalyzed an “evidence-based policy movement” in which many governments have implemented behavioral insight units or teams. From within these units, policy makers draw on scientific expertise about behavior to tackle complex societal issues such as energy usage, health decisions, and organ donor shortages.

The aim of this thesis is to investigate the evidence for behavioral policies. Such evidence for behavior change is subject to interpretation and debate. This thesis provides insights into how both the empirical evidence for behavioral policies (e.g., the experiments) and normative aspects of their interpretation (e.g., rationality evaluations) are generated and what this implies for evidence-based policymaking and policymakers. The central question is: on which underlying normative and empirical assumptions is the interpretation of evidence for behavioral policies based?

I focus on the role of agents’ perspectives in the interpretation of decision-making. In particular, I investigate to what extent behavioral evidence supports the idea that behavioral policies rectify the outcomes of people’s decision-making failures. The evidence may show that people change their behavior after seemingly irrelevant changes in their environment, but not why they do so. And the evidence may show that people behave irrationally, but only based on assumptions about people’s views of the choice options and according to very specific and highly debated theories about rationality. I argue that interpreting the existing experimental evidence in the right way is crucial to develop successful behavioral policies. Failing to consider the controversial aspects of behavioral evidence and expertise could lead to policymaking that is inefficient, ineffective and which may even backfire.

Chapter 1 introduces the thesis and provides a brief background of the historical context of research and theory development on decision-making and rationality. Furthermore, it discusses the concept of evidence-based policymaking, including a brief overview of some debates about the role of values and norms in science.

The remainder of the thesis consists of two parts, each containing one theoretically oriented chapter and one more practically oriented chapter. *Part 1* focusses on the interpretation of *behavior change* demonstrated in a wide variety of behavioral experiments. *Part 2* of the thesis analyzes the implications of the theoretical (dis)connections of behavioral policies for *evidence-based* policymaking and policymakers.

Part 1. Behavior Change

Chapter 2 shows that behavioral economics, and specifically “Nudge”, rely on a specific interpretation of the experimental results as irrational inconsistencies. I argue that this interpretation itself implicitly presupposes evidence for agents’ cognitive processes and beliefs about the choice options. Furthermore, I show that the normative role of agents’ cognitive processes and beliefs in the analysis of the decision problem is theoretically contested. This poses what I call the “Individuation Challenge” for behavioral policies. The challenge holds that for behavioral policies to be justified as evidence-based, an (explicit) theory of how options should be described for analysis should be developed. This includes the role of agents’ views on the choice options. Without such theory, behavioral policies cannot be accurately justified as helping irrationally inconsistent agents make better choices, as judged by their own lights.

Chapter 3, co-authored with Prof. Dr. N.J.A. van Exel, explores one way to gather more insight in agents’ own views regarding choice options is to extend quantitative research with qualitative insights. It uses a mixed-method study to gain insights into how people think about the choice to register as an organ donor. The study shows that agents’ views can generally be categorized into four different views. Two of them provide strong indications for why agents stick with defaults without being irrational, or so I argue.

Part 2. Evidence-Based Behavioral Policy

Chapter 4 explains the importance of understanding the (dis)connection between theories and policies in evidence-based behavioral policymaking. Based on Grüne-Yanoff and Hertwig (2016), I argue that the connection between the policies and their originating research programs is often weak, and sometimes even conflicting. I show that this is crucial in two ways: first, it is crucial for the evidence-based policy maker to realize this, as assuming theory-policy coherence while it does not exist or is weaker than assumed can lead to an incorrect assessment of evidence and ultimately to adoption of policies on false grounds. Second, assessing “theory-policy coherence” (TPC) may be helpful for the policy maker in the search and evaluation of further evidence. However, I argue that TPC cannot be the sole criterion for determining how well grounded a policy is in theory. It is too permissive because it may establish strong theory-policy coherence on vague theories. It may also be too restrictive if only applied to the policies themselves and not to assumptions that are crucial for comparing and selecting one policy over the other.

Chapter 5 communicates the evidential challenges identified in this thesis to the policymaker. It discusses what policymakers should do in response to the evidential challenges presented. I argue that policymakers should be aware of various behavioral policies, their alternatives, and scholarly disagreements associated with them, rather than focusing solely on the Heuristics and Biases research program and the behavioral policy of Nudge that originated from it. The chapter proposes the ‘3xE framework’ to expand and modify existing policy guidelines and tools, making the behavioral evidence and underlying values more explicit and open for debate. The framework includes using expert panels, literature reviews, evidence gap maps, and evaluating multiple policy alternatives.

Chapter 6 concludes and summarizes by discussing the evidential challenges and implications for evidence-based behavioral policies.

Nederlandse samenvatting

Recente ontdekkingen in de psychologie en de gedragseconomie tonen aan dat kleine, schijnbaar irrelevante, veranderingen in de manier waarop opties worden gepresenteerd grote effecten hebben op de keuzes die mensen maken. In deze context lijken mensen vaak irrationele besluitvormers te zijn. Deze zogenaamde "gedragsinzichten" uit de gedragseconomie en de psychologie, ook wel "behavioral insights" genoemd, hebben onderzoekers en beleidsmakers geïnspireerd tot het ontwikkelen van beleidsinterventies. Deze interventies zijn erop gericht zijn om mensen te helpen om betere beslissingen te nemen voor de samenleving en voor zichzelf. Kleinere bordjes leiden er bijvoorbeeld toe dat mensen zich beter aan hun dieet houden, door standaard "opt-out registratiesystemen" melden meer mensen zich aan als orgaandonor, en "smart meters" met informatie over het waterverbruik zorgen ervoor dat mensen minder lang douchen. Deze veranderingen in de keuzeomgeving zijn relatief eenvoudig en kostenefficiënt. Mede daardoor hebben deze gedragsinzichten zelfs een "evidence-based policy movement" gekatalyseerd. Dat wil zeggen, een beweging waarbij beleid vaker op wetenschap gebaseerd wordt. Zo hebben overheden steeds vaker speciale departementen voor gedragsbeleid. Binnen deze departementen gebruiken beleidsmakers wetenschappelijke expertise over gedrag om complexe maatschappelijke vraagstukken aan te pakken, zoals de energietransitie, preventieve gezondheid, en tekorten aan orgaandonoren.

Het doel van dit proefschrift is het onderzoeken van het wetenschappelijke bewijs waar gedragsbeleid op gebaseerd is. Dergelijk bewijs is complex en onderhevig aan interpretatie en debat. Het proefschrift geeft inzicht in hoe zowel het empirische bewijs voor gedragsbeleid (de experimenten), als de normatieve aspecten van de interpretatie ervan (bijv. irrationaliteit) tot stand komen en wat dit betekent voor "evidence-based policy" en beleidsmakers. De centrale vraag is: op welke onderliggende normatieve en empirische aannames is de interpretatie van bewijs voor gedragsbeleid gebaseerd?

Ik focus op de rol van het perspectief van mensen zelf in de interpretatie van hun besluitvorming. In het bijzonder onderzoek ik in hoeverre gedragsexperimenten het idee ondersteunen dat gedragsbeleid de uitkomsten van irrationele beslissingsprocessen van mensen corrigeert. De experimenten tonen vaak aan *dat* mensen hun gedrag veranderen na schijnbaar irrelevante veranderingen in hun omgeving, maar niet *waarom* ze dat doen. Bovendien kunnen experimenten aantonen dat mensen zich irrationeel gedragen, maar alleen op basis van veronderstellingen over hoe mensen zelf de keuzeopties zien en volgens zeer specifieke en sterk bediscussieerde theorieën over rationaliteit. Ik beargumenteer dat een accurate en genuanceerde interpretatie van het bestaande experimentele bewijs cruciaal is voor het ontwikkelen van succesvol gedragsbeleid. Het niet in overweging nemen van de controversiële aspecten van gedragsbewijs en gedragsexpertise kan leiden tot beleidsvorming die inefficiënt en ondoeltreffend is en kan zelfs averechts werken.

Hoofdstuk 1 introduceert het proefschrift en geeft een korte achtergrond van de historische context van onderzoek en theorieontwikkeling over besluitvorming en rationaliteit. Verder wordt het concept van evidence-based policy toegelicht, inclusief een kort overzicht van enkele debatten over de rol van waarden en normen in wetenschap en beleidsvorming.

De rest van het proefschrift bestaat uit twee delen, die elk een theoretisch georiënteerd hoofdstuk en een meer praktijkgericht hoofdstuk bevatten. *Deel 1* richt zich op de interpretatie van gedragsverandering die is aangetoond in verschillende gedragsexperimenten. *Deel 2* van het proefschrift analyseert de implicaties van de theoretische (dis)connectie van gedragsbeleid voor evidence-based policy en beleidsmakers.

Deel 1. Gedragsverandering

Hoofdstuk 2 laat zien dat de gedragseconomie, en in het bijzonder “Nudge beleid”, uitgaan van een specifieke interpretatie van de experimentele resultaten. Volgens deze interpretatie vertonen mensen die hun keuze wijzigen naar aanleiding van een kleine verandering van de presentatie van die keuzes “irrationeel inconsistent gedrag”. Ik betoog dat deze interpretatie impliciet vooronderstellingen bevat over mensen hun cognitieve processen en hoe ze die keuzes interpreteren. Voor deze vooronderstellingen is weinig bewijs en zonder deze vooronderstellingen kan niet gesteld worden dat de beslissingen inconsistent zijn. Daarnaast laat ik zien dat de normatieve rol van de cognitieve processen en overtuigingen van mensen in de analyse van de gedragsverandering überhaupt controversieel is in de wetenschappelijke theorie. Ik beargumenteer dat dit een uitdaging vormt voor gedragsbeleid. De uitdaging houdt in dat een (expliciete) theorie moet worden ontwikkeld over hoe opties moeten worden ‘geïndividueerd’ in de gedragsanalyse. Dat wil zeggen, dat een theorie moet worden ontwikkeld over welke aspecten van de keuze-opties moeten worden meegenomen wanneer men een keuze analyseert. Ik beargumenteer dat dit de rol van de opvattingen van mensen zelf over de keuzeopties die aan hen gepresenteerd worden moet omvatten. Zonder een dergelijke theorie kan niet bepleit worden dat gedragsinterventies op wetenschap gebaseerd beleid is dat mensen helpt om betere beslissingen te nemen.

Hoofdstuk 3, geschreven samen met Prof. Dr. N.J.A. van Exel, verkent een manier om meer inzicht te krijgen in de eigen opvattingen van agenten over keuzeopties door kwantitatief onderzoek uit te breiden met kwalitatieve inzichten. Het gebruikt een “mixed-method” studie om inzicht te krijgen in hoe mensen denken over de keuze om zich te laten registreren als orgaandonor. Uit de studie blijkt dat de opvattingen van mensen over het algemeen kunnen worden ingedeeld in vier verschillende opvattingen. Twee daarvan geven sterke indicaties waarom agenten vasthouden aan “defaults” zonder irrationeel te zijn.

Deel 2. Evidence-based gedragsbeleid

Hoofdstuk 4 onderzoekt het belang van coherentie tussen theorie en gedragsbeleid (“Theory Policy Coherence”) voor evidence-based policy. Op basis van Grüne-Yanoff en Hertwig (2016) betoog ik dat de verbinding tussen het beleid en hun oorspronkelijke onderzoeksprogramma's vaak zwak en soms zelfs tegenstrijdig is. Ik laat zien dat dit op twee manieren cruciaal is: ten eerste is het cruciaal voor de wetenschap-gedreven beleidsmaker om zich dit te realiseren, omdat het aannemen van samenhang tussen theorie en beleid, terwijl deze niet bestaat of zwakker is dan verondersteld, kan leiden tot een onjuiste beoordeling van het bewijs en uiteindelijk tot het aannemen van beleid op verkeerde gronden. Ten tweede kan het beoordelen van samenhang tussen theorie en beleid nuttig zijn voor de beleidsmaker bij het zoeken naar en evalueren van verder bewijsmateriaal. Ik betoog echter ook dat samenhang tussen theorie en beleid niet het enige criterium kan zijn om de wetenschappelijke (theoretische) basis van beleid te beoordelen. Aan de ene kant is het concept te

breed, omdat het toelaat dat beleid als goed gegrond in de theorie wordt beoordeeld in het geval dat de wetenschappelijke theorie te vaag is. Aan de andere kant kan het ook te restrictief zijn als alleen de samenhang van de wetenschap met het beleid zelf onderzocht wordt. Dit speelt bijvoorbeeld wanneer alleen gekeken wordt naar samenhang van de theorie met het beleid zelf, maar niet naar andere veronderstellingen die meespelen bij het vergelijken en afwegen van verschillende beleidsinterventies. Dat zijn bijvoorbeeld veronderstellingen dat een bepaald type beleid beter is dan een ander type beleid, omdat het mensen zou leren om betere beslissingen te nemen. Ook deze typen veronderstellingen moeten beoordeeld worden op hun wetenschappelijke basis, beargumenteer ik.

In *hoofdstuk 5* worden de in dit proefschrift vastgestelde uitdagingen voor het wetenschappelijke bewijs voor gedragsbeleid voor beleidsmakers beschreven. Het hoofdstuk bespreekt wat beleidsmakers moeten doen als reactie op de gepresenteerde uitdagingen met betrekking tot bewijs voor gedragsbeleid. Ik betoog dat beleidsmakers zich bewust moeten zijn van verschillende mogelijke gedragsinterventies, hun alternatieven en de daarmee samenhangende wetenschappelijke basis en twistpunten. Het hoofdstuk presenteert het '3xE kader' om bestaande beleidsrichtlijnen en -instrumenten uit te breiden en aan te passen, door het gedragsbewijs en de onderliggende waarden explicieter te maken en open te stellen voor debat. Het kader omvat het gebruik van deskundigenpanels, literatuuronderzoek, het in kaart brengen van leemtes in het bewijsmateriaal en het evalueren van meerdere beleidsalternatieven.

Hoofdstuk 6 concludeert en vat samen door de uitdagingen en implicaties voor wetenschap-gedreven gedragsbeleid te bespreken.

