

Experiences and Perceptions of Default Nudges

Experiences and Perceptions of Default Nudges

Patrik Michaelsen

Doctoral Dissertation in Psychology
Department of Psychology
University of Gothenburg
September 10, 2021

© Patrik Michaelsen

Cover layout: Sofia Calderon

Printing: Stema Specialtryck AB, Borås, Sweden, 2021

ISBN: 978-91-8009-447-4 (PDF)

ISBN: 978-91-8009-446-7 (Print)

ISSN: ISSN 1101-718X Avhandling/Göteborgs universitet, Psykologiska inst.

<http://hdl.handle.net/2077/69058>



For Beda

Abstract

Of the many challenges we face as individuals and as a society, several of the most important and hardest ones are behavioral. We often know what needs to be done to solve an issue, but struggle to put this knowledge into practice. It is the case for tackling global warming, just as it is for tackling the COVID-19 pandemic.

Insights from psychology and other behavioral sciences are increasingly often used by policymakers to design interventions that promote behavior change in domains such as health, personal finance, and environmental protection. The approach is sometimes referred to as “nudging”, likening the intervention to a gentle nudge in the desired direction. Proponents of nudging maintain that nudges are cost-effective policy instruments that respect people’s freedom of choice. Critics, however, argue that the approach fails on both grounds.

In this thesis, I argue that how people perceive and experience nudge interventions is an underappreciated factor that can help assess, and potentially address, both effectiveness and ethical concerns. In the introduction, I outline a framework for how this can be the case. Specifically, I propose that people’s perceptions and experiences are relevant to assessing (1) the ethics of nudging, (2) the net effect of nudging, and that they may (3) be moderators of the success of nudges influencing behavior. I then present three empirical studies (nine sub-studies, total $N = 5171$) that used online experiments to assess how people perceive and experience being subjected to nudges. In contrast to the majority of similar research, the present studies primarily subjected participants to nudges first-hand. That is, participants did not rate descriptions of nudges, but engaged in choice tasks first hand before assessing the intervention. Results indicated that people subjected to default nudges: experienced themselves as highly, and not less, autonomous and satisfied with their choices (Study I); found the use of the nudge to be fair, though less fair than no nudge (Study II); and favorably perceived a choice architect using the nudge, and not less so than in the absence of a nudge (Study III).

Additionally, the studies found that making the nudge increasingly transparent by disclosing its presence and expected effect to participants did not meaningfully change their experiences of choosing, but could either improve or worsen perceptions of the nudge depending on the circumstances of the evaluation. When participants were disclosed of the nudge after they had already made a choice, this was found to negatively affect their perceptions of the choice architect. Providing a nudge disclosure did not, however, influence participants’ propensity to act in line with the nudge.

I conclude that the present findings paint a generally positive picture of how default nudges are perceived and experienced, but that more research is necessary to properly inform policy. I suggest that policymakers should routinely use measures of choice experiences as a guide when designing new nudge interventions. Nevertheless, as judged by the empirical evidence available at present, default nudges appear to be a viable form of intervention in relation to concerns about both their effectiveness and ethicality.

Swedish summary

Denna doktorsavhandling undersöker personers upplevelser av och omdömen om nudge-interventioner. Nudging är ett paraplybegrepp för beteendebaserade interventioner som fungerar genom ändringar i beslutsmiljön ("beslutsarkitekturen") en individ ställs inför. Genom att använda insikter från psykologisk forskning om vad som påverkar människors beslutsfattande är det möjligt att utforma beslutsmiljöer så att vad som menas vara de mest fördelaktiga valalternativen är mer sannolika att bli valda. Nudge-interventioner kan användas för att påverka beteenden i många olika situationer och avseenden, som exempelvis för att uppmuntra människor till mer hälsosamma, miljövänliga och finansiellt hållbara val. Ett exempel på en nudge kan vara att omorganisera placeringen av hälsosamma varor i ett café eller en affär så att de hälsosamma varorna är mer synliga än mindre hälsosamma alternativ. Eftersom det då är enklare att lägga märke till den hälsosamma varan blir det mer troligt att det är den vi köper, och det kan sägas att vi genom ändringen av placeringen då har blivit nudgade till ett mer hälsosamt val. Ett annat exempel kan vara att informera om en social norm, exempelvis att andra besökare till parken vi befinner oss i inte lämnar något skräp på marken efter sig. Eftersom vi vanligen dras till att vilja göra likadant som andra människor kan detta normmeddelande påverka oss till att låta bli att skräpa ner. Ett tredje exempel, vilket är den typ av nudge-intervention som det kommer att fokuseras på i denna avhandling, är att använda sig av den så kallade förvalt-effekten. Med förvalt-effekten menas att i valsituationer där vi ställs inför ett flertal alternativ så är vi ofta i högre grad benägna att välja ett visst alternativ när detta är förvalt, "för-ikryssat", och vi aktivt måste ändra till ett annat alternativ för att byta. Förvalt-effekten kan på så vis användas som en nudge genom att det valalternativ som eftersträvas sätts som förvalt. Exempelvis kan en affär sätta miljövänlig frakt som det förvalda fraktalternativet när man beställer från deras onlinebutik.

En del kritiker har dock opponerat sig mot användandet av nudge-interventioner. De har ifrågasatt, exempelvis, om interventionerna är effektiva nog att åstadkomma nödvändiga beteendeförändringar, samt om det är etiskt att försöka påverka andra med hjälp av denna typ av "psykologiska knep". I denna avhandling argumenterar jag för att forskning om personers upplevelser av och omdömen om nudge-interventioner kan hjälpa oss att utvärdera skäligheten i denna kritik.

I avhandlingens första del utvecklar jag ett ramverk i vilket jag föreslår att personers upplevelser av och omdömen om nudge-interventioner är viktiga att beakta i åtminstone tre bemärkelser. En första aspekt är att de kan informera den etiska debatten som pågår kring nudging. Detta kan ske dels genom att informera om allmän opinion, och dels genom att testa hur väl teoretiska antaganden som förekommer i normativa debatter håller empiriskt. En andra aspekt är att upplevelser och omdömen från personer utsatta för nudge-interventioner bör räknas mot interventionens totaleffekt. Exempelvis kan nudge-interventioner sannolikt oftare än mer traditionella policy-interventioner göra en valsituation roligare eller enklare att ta ställning till för den som väljer. Omvänt kan risken finnas att vissa nudge-interventioner gör personer

frustrerade, exempelvis om interventionen upplevs manipulativ. I dessa fall är totaleffekten för nudge-interventionen större respektive lägre än vad som mäts enbart i form av beteendeförändringen. En tredje aspekt är att upplevelser av och omdömen om nudge-interventioner kan tänkas påverka hur framgångsrik interventionen är i att åstadkomma beteendeförändring, både i stunden och i framtida interaktioner mellan den som använt och den som blivit utsatt för nudge-interventionen. Exempelvis skulle en person som upplevt interventionen positivt kunna bli mer benägen att bifalla avsändaren av nudge-interventionen i framtiden (och det omvända). Förutom upplevelser av och omdömen om nudge-interventioner undersöker avhandlingen också hur personer reagerar på att bli upplysta om en nudge-intervention, att den är på plats i valsituationen och hur den kan förväntas påverka personers beslut, i samband med valsituationen.

Tre artiklar med totalt nio delstudier och 5171 försöksdeltagare undersökte dessa teman. Samtliga studier fokuserade på nudging med hjälp av den så kallade förvalt-effekten. På grund av sin relativt stora effekt på beteende, och på grund av att förvaltnudgar ofta anses svåra att upptäcka –och därför potentiellt manipulativa–, är de extra intressanta att studera i relation till vad personer anser om användandet av dem och hur de upplever det att bli utsatta för dem. Förvalt-nudgen är dessutom en av de vanligast använda och beforskade typerna av nudge-interventioner vilket gör avhandlingens resultat enklare att generalisera till andra situationer.

Studie I fokuserade på huruvida att bli utsatt för en förvalt-nudge påverkade deltagares upplevelse av valet, deras upplevda autonomi och nöjdhet med valet, samt huruvida de ansåg att nudgen hotade deras frihet att välja fritt och om de invände mot användandet av den. I tre experiment utsattes deltagare för förvalts-nudgar i miljörelaterade beslutssituationer, och deras upplevelser jämfördes med deltagare som ställdes inför samma valsituation men utan någon förvalt-nudge. I två av experimenten manipulerades även interventionens nivå av transparens genom att hälften av deltagarna blev presenterade med en textruta innan sitt val som informerade dem att valsituationen var utformat på ett vis som kunde påverka deras beslut och i vilken riktning påverkan kunde förväntas ske. Det vill säga, i förvalt-nudgegruppen fick deltagarna veta att valsituationen var utformad med det prosociala valalternativet förvalt och att detta kunde leda till att de blev mer sannolika att välja det prosociala alternativet. I vart och ett av experimenten hittades en stark påverkan av förvalt-nudgen på deltagarnas val. Det fanns däremot inget stöd för att deltagarna i delstudierna upplevde sig mindre i kontroll av sitt beslutsfattande (mindre autonoma) eller mindre nöjda med sina val som ett resultat av att ha blivit utsatta för förvalt-nudgen. Dock fanns det en viss tendens att se nudgen som ett inkräktande på ens frihet att välja fritt, men skattningarna var ihållande i den lägsta tredjedelen av skalan. Personer som blivit utsatta för förvalt-nudgen invände inte i högre grad mot hur valsituationen var utformad än vad andra deltagare gjorde. Att ha blivit informerad eller inte blivit informerad om förvalt-nudgen påverkade inte i vilken grad deltagarna valde i linje med den. Det hade heller nästan inga effekter på deltagarnas upplevelser av eller omdömen om interventionen; det enda statistiskt signifikanta resultatet var en något högre perception av att nudgen försökte inkräkta på deltagarens frihet att välja fritt.

Studie II fokuserade på hur etiska ("fair") deltagare fann förvalt-nudges, och huruvida att bli informerad om interventionen i förväg påverkade denna bedömning. I tre delstudier varierade vi huruvida deltagarna fick bedöma en beskrivning av en förvalt-nudge eller bli utsatt för den själv, samt huruvida en jämförelsepunkt för förvalt-nudgen gavs. I samtliga delstudier rörde valsituation ett beslut mellan att skänka en mindre summa pengar (20€) till välgörenhet (en barncancerfond) eller behålla pengarna själv. I de versioner av valsituationen som hade en förvalt-nudge var det förvalda alternativet att skänka pengarna till välgörenhetsorganisationen. Resultaten visade att deltagarna genomgående skattade valsituationens utformning som mest etisk när det inte fanns någon förvalt-nudge. Att göra förvalt-nudgen transparent genom att informera beslutsfattare om den i förväg ledde till högre skattningar när deltagarna fick jämföra en transparent och en icke-transparent version sida vid sida. Utan möjligheten att jämföra försvann dock övertaget för den transparenta versionen, och när deltagare upplevde förvalt-nudgen själva, dvs. själva fattade ett beslut i valsituationen innan de bedömde den, skattades den transparenta versionen som mindre etisk än den icke-transparenta versionen av förvalt-nudgen. I samtliga delstudier skattades dock både den transparenta och icke-transparenta versionen av förvalt-nudgen på den övre, positiva, delen av skalan. Sammantaget visade Studie II att ökad transparens inte är en enkel lösning på hur förvalt-nudgear kan presenteras på ett sätt som framstår som mer etiskt, utan risken finns att tilltaget har negativa konsekvenser i praktiken.

Studie III fokuserade på hur personer utsatta för en förvalt-nudge bedömde avsändaren av nudgen, och hur denna bedömning påverkades av att deltagaren blev informerad om nudgen. I två delstudier varierade vi huruvida deltagarna blev informerade om nudgen före eller efter att de hade fattat sitt beslut i experimentets valsituation. Valsituationen handlade om att tacka ja eller nej till att oavlönat fylla i en kort extra enkät till ett studentarbete efter att huvudenkäten var slut. För deltagare som hamnade i en experimentgrupp med en förvalt-nudge var det förvalda alternativet att tacka ja till att fylla i den extra enkäten. I den första delstudien blev en grupp deltagare informerade om förvalt-nudgen innan de gjorde sitt val, och svarade därefter på frågor om hur de bedömde avsändaren av nudgen (exempelvis hur pålitlig, öppen och manipulativ denne framstod). En annan grupp blev inte informerad om nudgen innan de gjorde samma val och svarade på samma frågor om avsändaren. Därefter blev båda grupperna informerade om nudgen (dvs. första gruppen blev påmind och den andra gruppen blev informerad för första gången) och fick sen möjlighet att ändra sitt val och sin bedömning av avsändaren. Vi fann att vid första mätningen var det ingen skillnad mellan grupperna sett i deras val eller i hur de bedömde avsändaren av nudgen. Vid den andra mätningen gavs dock en mindre positiv bedömning av avsändaren, framför allt i gruppen som blivit informerad i efterhand. I ingen av grupperna ändrade deltagarna dock nämnvärt på sitt beslut om de kunde tänka sig att fylla i den extra enkäten. Den andra delstudien fokuserade på huruvida att bli informerad om nudgen, samt deltagarnas bedömning av avsändaren, påverkade deltagarnas beslut i en andra, separat, valsituation (som inte var utsatt för någon nudge). I den andra valsituationen frågades deltagarna om de kunde tänka sig att skriva upp sin e-postadress för att längre fram bli kontaktade angående oavlönat deltagande i en annan studie. Likt i den första

delstudien fann vi att deltagare som blivit informerade i efterhand om nudgen i det första valet bedömde avsändaren mindre positivt. Vi hittade dock inget samband mellan denna minskning och i vilken grad deltagarna tackade ja i det andra valet. Deltagarnas benägenhet att ställa upp i det andra valet var lika stor oavsett om de blivit utsatta för en förvalt-nudge eller inte i det första valet, och oavsett om de blivit informerade om nudgen i förväg, i efterhand eller inte alls. Sammantaget indikerade Studie III att personer som använder förvalt-nudgear verkar kunna undgå att bli sämre ansedda genom att vara transparenta med att de använder en nudge. Inget stöd hittades dock för att personerna som blivit utsatta för nudgen reagerade tillräckligt negativt att detta ändrade deras beslut i efterkommande interaktioner med samma avsändare av nudgen.

Sammantaget fann de tre studierna att förvalt-nudgear uppfattades relativt väl. I relation till etiska frågor så indikerade resultaten att när man ser till personers upplevelser av och omdömen om nudge-interventioner så var dessa mer positiva än vad kritiken gjort gällande. Centralt så upplevde inte deltagare som blivit utsatta för en nudge att deras autonomi inskränktes på. I relation till att upplevelser och omdömen om nudge-interventionen kan påverka interventionens totaleffekt hittades varken större positiva eller negativa resultat. Exempelvis hittades ingen skillnad i nöjdhet med ens val mellan deltagare som blivit utsatta för en förvalt-nudge och inte. Förutom detta indikerade studierna att en förvalt-nudge inte blir mindre effektiv av att personer informeras om interventionen i förväg. Beroende på när informationen ges och under vilka omständigheter bedömningen görs kan dock ökad transparens leda till antingen bättre eller sämre omdömen om nudgen och dess avsändare.

List of Empirical Works

This thesis consists of an introductory chapter and the following three studies, which are referred to by their Roman numerals:

- I. Michaelsen, P., Johansson, L-O., & Hedesström, M. (2021). Experiencing default nudges: Autonomy, manipulation, and choice-satisfaction as judged by people themselves. *Behavioural Public Policy*. Advance online publication. doi:10.1017/bpp.2021.5
- II. Michaelsen, P., Nyström, L., Luke, T. J., Johansson, L-O., & Hedesström (2021). Are default nudges deemed fairer when they are more transparent? People's judgments depend on the circumstance of the evaluation. *Manuscript submitted for publication*.
- III. Michaelsen, P., Nyström, L., Luke, T. J., & Hedesström, M. (2021). Downstream consequences of disclosing defaults: Influences on perceptions of choice architects and subsequent behavior. *Manuscript granted in-principle acceptance by Comprehensive Results in Social Psychology*.

Contents

Acknowledgements	iii
Preface	v
Introduction	1
Challenges for nudging.....	5
Effectiveness issues	6
Ethical issues	7
People’s perceptions and experiences of nudges as partial solutions: A framework	10
Principle 1: People’s perceptions and experiences inform the ethical assessment of nudges...	11
Principle 2: People’s perceptions and experiences are part of the net effect of the nudge	13
Principle 3: People’s perceptions and experiences can moderate behavior change.....	14
The relevance of intervention transparency	14
How do people perceive and experience nudges?.....	15
Description-based research	16
Experience-based research.....	19
Transparency and choice	23
Summarizing the state of knowledge.....	24
Summary of Empirical Studies	29
Study I	31
Study 1: Experiencing a pro-environmental default nudge	32
Study 2: Do choice experiences deteriorate when intervention transparency is increased?	33
Study 3: Do choice experiences deteriorate when choice stakes are increased?	34
Study 4: Internal meta-analysis and equivalence tests	35
Study II.....	36
Study 1: Description-based joint evaluation	37
Study 2: Description-based separate evaluation.....	38
Study 3: Experience-based separate evaluation	39
Study III.....	40
Study 1: Perceptions of the choice architect and timing of nudge disclosure.....	41
Study 2: Perceptions of the choice architect and subsequent behavior	43
General Discussion	45
Taking stock of people’s perceptions and experiences of default nudges	45
Methodological limitations and considerations.....	49
On theory and practical implications	54
Concluding remarks.....	59
References	61

Acknowledgements

I began my Ph.D. studies in 2016. At this time, the recruitment process at the Department of Psychology still consisted of an open call to which hopeful prospective students sent in research plans for review. Sure, a few months before my application was posted I was vaguely familiar with the concept of nudging; you know, I had heard the name. Yet it is absolutely certain that no research idea I could have dreamed up on the topic would have had any chance whatsoever of getting me accepted into the program. So thank you, very much, Martin Hedesström and Lars-Olof Johansson for –without really knowing me– believing in me enough that you let me write an application based on your research idea (and to Tommy Gärling who I assume put in a good word). And thank you for believing sufficiently in me that you subsequently let me pursue my own research ideas once they arose. The ideas were not always directly on the original topic, but I am happy to note that in the end my thesis project came full circle and started (directly) investigating issues promised in that research plan. I am very grateful that you let me think and work as independently as you did.

Martin, thank you in particular for your meticulous feedback on everything I have done. Very few people operate at the level where they can spot a bold-faced blank space. If my thesis ended up readable, it is in great part thanks to your generous and careful commenting on my texts through these years. Lars-Olof, thank you for mentoring me on the opposite: that one must transcend getting lost in the details if one wants to get a message across to others (as well as to get anything done). You have the personality of a Telecaster even though you keep insisting on playing the Strat. That is obviously a compliment. You two complement each very well and it has been a privilege to be your supervisee.

There was a divide about halfway through my Ph.D. studies, where in the latter half first Lina Nyström and later Timothy Luke joined the endeavor. That latter half had all the good parts of the first (and more), just in even better renditions. That was thanks to your company: thank you! Lina, during this time you were the most capable research assistant thinkable and frequently made me wonder when Martin would realize and switch our contracts. Timothy, you are the most comprehensively skilled academic I have met. That is great and all on its own, but coupled with a Jonas Burén-like generosity and helpfulness I do not know what more one could ask for in a collaborator. Yet on top of your (both of you) intellectual qualities, thank you most of all for being (at least) equal parts silly.

Speaking of generosity, thank you Tommy Gärling (once more) and Amelie Gamble for mentoring me before I started as a Ph.D. student (as well as a long time into it). Without your early support and encouragement I would have been someplace else today. I am not sure where, but I am certain that place would have been lesser.

Several people have provided great help towards the end of the the completion of this thesis. First and foremost I would like to thank Ann Backlund for guiding me through the process, for making sure all of the many tasks got carried out correctly, and for looking out for me. Next, I am very thankful to the persons, apart from my supervisors, who have reviewed and commented on earlier drafts of this thesis: thank you Per Kristensson, Carl Martin Allwood and Tommy Gärling (yet once more). Last, I am also incredibly thankful to the people who agreed to be a part of the defense of the thesis: Liam Delaney, who will be acting as the external opponent, Erik Angner, Sandra Buratti, and Per Kristensson (once more), who will make up the examining committee, and to Karl Ask, who will be acting as the chairperson. I am truly honored by each one of you agreeing to take part.

So many colleagues made the time as a Ph.D. student stimulating and enjoyable. Apart from those already mentioned, thank you Magnus Bergquist, Cecilia Bergstad, Pär Bjälkebring, Jonas Burén, Gró Einarsdottir, Py Eriksson, Emelie Ernberg, Gustaf Glavå, Maria Gröndal, Fanny Gyberg, André Hansla, Isabelle Hansson, Malin Joleby, Lukas Jonsson, Hanna Larsson, Debora Lombardi, Erik Mac Giolla, Mikaela Magnusson, Elaine McHugh, David Neequaye, Andreas Nilsson, Érika Ramos, Emma Renström, Isac Sehlstedt, Johan Skoog, Meghana Srivastav, Leif Strömwall, Erik Sturén, Andrea Valik, Lina Wirehag Nordh, and Linn Zulka. In particular, thank you Sofia Calderon and Emma Ejelöv for always being there.

A few non-sentient entities also deserve a mention. Thank you to the departmental research groups I have belonged to, JDMS and EPU, for being stimulating and supporting. Thank you CeCAR for creating and maintaining a generous and inspiring interdisciplinary research environment. Thank you to the Society for Judgment and Decision Making and to Open Science Community Sweden for providing intellectual homes. On another level, thank you Kathmandu restaurant, Delitalia, Hagabion, Stigbergets bryggeri, Kaffelabbet, Tikket cycling (non-)club, and cycling in general.

Finally, there is one (sentient being) who outmatches everyone and everything mentioned above: Ellen. You have supported me in more ways than I could summarize in this space. Thank you. I love you.

Patrik
Gothenburg, August 2021

Preface

The aim of this thesis is to argue for the value of people's perceptions and experiences in assessments of nudge interventions, and to investigate a number of central such perceptions and experiences empirically. A subplot of the thesis concerns how choices, perceptions, and experiences are affected by efforts to make a nudge less manipulative through increasing the transparency of its presence and anticipated influence.

The topic is motivated by the recent proliferation of policy interventions informed by behavioral science around the world, and more specifically by the criticisms this approach to behavior change has drawn. In particular, critics of a nudge approach to policy interventions have argued that this type of interventions can fail ethical standards. If accepted, the criticism strikes a devastating blow to the enterprise. In this thesis I argue that properly appreciating people's perceptions and experiences of nudge interventions is a meaningful way of assessing, and potentially addressing, these concerns.

The contents of the thesis are structured in the following way. The first half provides a background to the topic. It begins with a brief introduction of nudging and has a focus on so-called default nudges, which is the type of nudge researched in the empirical work later on. Next follows a section where the two major lines of criticism nudges have drawn are outlined, criticisms of their effectiveness and ethicality. After the criticisms have been described, I develop a framework for how people's perceptions and experiences of nudge interventions can be used to assess the merits of the criticisms, and potentially serve as partial solutions to them. I thereafter review the previous literature on the topic and conclude that important issues remain to be studied.

The second half of the thesis consists of the empirical work. A brief summary introduces the specific topics of each of the three studies and connects them to the framework outlined in the first half of the thesis. Next, the methods and major results from each study are presented. The full journal article-versions of each study are presented in appendices found at the end. The thesis concludes with a synthesis of the empirical studies and a general discussion of their findings, strengths and limitations.

Introduction

How to succeed in making good decisions is truly an age-old question. As decision makers, people struggle with limited knowledge, unclear and unstable preferences, behavioral inertia, and procrastination, to name but a few obstacles. As members of communities and societies, our struggles are not uniquely our own, but also affect, and are affected by, those around us. Because we do not live in a vacuum, others will want to influence our decision-making and behavior. On top of the persuasion attempts and stunts we might experience from peers, more structured and systematic attempts of organizations and governments will target us. Such behavior change interventions can come in many forms, and one way to organize them is on a continuum ranging from soft to hard interference with people's freedom of choice.

At the soft end, we find interventions such as providing people with new information. Learning new information will often lead us to update prior beliefs, which in turn can make us change behavior. For instance, if I want to become taller and hold the erroneous belief that eating lots of carrots will make me so, learning that this causal relationship does not hold will likely make me change my diet. This is the rationale for using information campaigns to change people's behavior: "if people just knew better, they would also do better". An information campaign is labeled a "soft" intervention because it does not limit the decision maker's freedom of choice. This is appealing from an ethical and policy perspective, as the autonomy of the decision maker is respected. A common criticism, however, is that as targeted interventions, information campaigns are largely ineffective at changing people's behavior (e.g., Conly, 2013).

At the hard end of the interference continuum, we find interventions such as mandates and bans. Mandates and bans are typically effective at changing behavior. For instance, a Swedish report on seat belt use in cars found that in 1983, when front seat belt use was mandatory but rear seat belt use was not, front seat riders were more than ten times likelier to use seat belts than were rear seat riders (Lacko & Nilsson, 1988). When rear seat belt use was mandated in legislation a few years later, rear seat belt use more than doubled in the week after the law was implemented. However, what may be gained in effectiveness

is lost in respect for the decision maker's autonomy, as making certain acts punishable is a clear restriction on freedom of choice.

Is it possible to have behavior change interventions that are both effective and freedom preserving? In 2008, the concept of *nudging* was launched as a policy initiative aiming to meet both ends (Thaler & Sunstein, 2008; but see also Camerer, Issacharoff, Loewenstein, O'Donoghue, & Rabin, 2003; Sunstein & Thaler, 2003). A "nudge" is characterized by its coiners as a strategic, gentle push that steers a decision maker in a desired direction. It is an umbrella term for behavior change interventions that build on insights into human decision-making and uses these insights to design choice environments in ways that make desired outcomes more probable. For instance, research shows that people are more likely to take a flu shot if the appointment is set up by default (Chapman, Li, Colby, & Yon, 2010). Thus, pre-booking people for appointments, which they are fully free to opt out of, can be used to nudge people into getting vaccinated. Chapman et al. (2010) found that when a flu shot appointment was booked by default, 45% showed up for the appointment and got vaccinated. In the control group, whose members were given the same information about the vaccination but had to book their appointments themselves, 33% showed up for an appointment. Encouraging a behavior by using this mechanism is an example of a specific nudge referred to as a *default nudge*. More broadly, in Thaler and Sunstein's (2008, p. 6) own words, a nudge is:

Any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting the fruit at eye level counts as a nudge. Banning junk food does not.

This definition, although not an attempt at being a formal one, is notably broad. In later writing, Thaler and Sunstein have often preferred to unpack the concept by providing examples of interventions instead of working out an exhaustive definition (e.g., Sunstein, 2018; Sunstein & Reisch, 2019). Despite subsequent definition attempts by other authors (e.g., Guldborg Hansen, 2016; Löfgren & Nordblom, 2020), a generally agreed on, exhaustive definition is still lacking. This conceptual ambiguity is likely a key factor behind much confusion and disagreement surrounding nudging. What have been perceived as "misconceptions" about nudges have been prevalent enough that one of the original authors even published an article aiming to debunk more than half a dozen of them (Sunstein, 2018).

Satisfactory or not, Thaler and Sunstein's (2008) definition does highlight two central features of nudging: 1) it has to do with the design of the "choice architecture"; and 2) it rules out hard restrictions on freedom of choice, for example, putting costs on alternatives or completely removing them.

The term "choice architecture" captures much of the essence of nudging. It refers to how environmental cues are shaped in situations in which we make decisions. Subtle cues can often influence how we process information and make decisions. An intentional arrangement of such cues in order to bring about a desirable outcome is paradigmatically a nudge.

The already mentioned default nudge is a prototypical example. A default nudge is an intervention that works by designating one course of action as the default, which will happen unless the decision maker actively switches away from it. For instance, imagine being at the checkout counter when making a purchase online. In most cases, there will be a choice between environmentally friendly or standard methods of shipping. If the seller wants to encourage environmentally friendly choices, they can make the green option the default course of action by preselecting it on the screen. This way, if customers do not make the active switch to standard shipping, the green shipping option will be chosen. The intervention can be described as the choice architect, that is, the person administering the nudge, formatting the choice architecture in an opt-out instead of an opt-in choice format. If people do not desire the preselected course of action, they will need to actively opt out of it. In an opt-in format, people would correspondingly need to actively change to the desired behavior, to opt in, for it to happen.

Preselecting options in this way is a versatile tool for behavior change and can be applied in many contexts, both small and large. For instance, changing the default setting when printing or copying documents from the one-sided to two-sided option has been found to reduce paper use (Egebark & Ekström, 2016). When an electricity provider supplies green instead of grey electricity by default, most people stick with the green electricity plan (Pichert & Katsikopoulos, 2008). In countries where being enlisted as an organ donor by birth is the default, upwards of 99% of the population remain enlisted; in contrast, when not being enlisted is the default, the proportion is in the 4–30% region (Johnson & Goldstein, 2003).

Defaults are thought to be effective by tapping into at least three different psychological mechanisms (Dinner, Johnson, Goldstein, & Liu, 2011; Jachimowicz, Duncan, Weber, & Johnson, 2019). One is that following a default saves the decision maker effort, in both a physical and psychological sense. By following a default, the decision maker avoids having to perform the action needed to opt out. In many cases this effort may be as little as checking a box in a form or making a few clicks on a computer screen, but other cases

may require more effort. For instance, obtaining, filling in, and mailing a form to a company or governmental agency. While still relatively easy to perform, the need to engage in such behaviors can be a sufficient deterrent. Following a default also means we can avoid having to engage with the decision intellectually. For instance, organ donation can be a complex issue to consider, so sticking with the default means one does not have to.

A second mechanism by which defaults may work is that the preselection of an option can be interpreted as a recommendation or implied endorsement (McKenzie, Liersch, & Finkelstein, 2006). In a situation in which the decision maker lacks knowledge or strong antecedent preference, an option preselected by a trusted choice architect may appear safe and sound. For instance, sticking with the default settings (vs. engaging with advanced options) when installing computer software may typically be driven by this mechanism.

Third, defaults can work by establishing a reference point for behavior (Dinner et al., 2011; Kahneman & Tversky, 1979). When the default is perceived as a reference point, thoughts of deviation can highlight the risks associated with change, in turn invoking loss aversion and making sticking with the default more likely. Keeping an insurance policy after a free trial period, even though one might not have chosen it in the first place, may be an example of this mechanism in effect.

The proposed mechanisms should not be viewed as mutually exclusive; rather, they may cooperate in any given instance. In the insurance example, for instance, unwillingness to make the effort needed to cancel the subscription may be an equally contributing factor.

Across a wide range of applications, default nudges display relatively potent psychological effects. In comparison with other types of nudges, defaults tend to have bigger impacts (Hummel & Maedche, 2019), and a meta-analysis of 58 studies found an average effect size corresponding to a Cohen's *d* of 0.68 (Jachimowicz et al., 2019).

Changing the decision frame from opt-in to opt-out is one example of how altering a choice architecture can strategically influence decisions. Several other techniques to nudge people have been documented in overviews and taxonomies (e.g., Hummel & Maedche, 2019; Johnson et al., 2012; Münscher, Vetter, & Scheuerle, 2016; Szaszi, Palinkas, Palfi, Szollosi, & Aczel, 2018). For instance, making select information salient, simplifying or translating information, grouping choice options, and providing social reference points are other subtle and easy-to-change factors that can be used to nudge people. The list of ways to nudge people is essentially endless, as it is limited only by our understanding of what affects decision making and by the inventiveness of the world's choice architects. Part of why it is so difficult to define the concept of nudging is thus that it includes such a wide range of types of interventions. Due

to this umbrella term status and lack of concrete unifying theory, nudging may be better viewed as a toolbox or kind of approach, rather than as a set list of solutions to behavioral problems. A common way of circumventing definitional issues, and broadening the scope, is to talk about using insights from behavioral science to inform behavior change interventions or other policy, rather than only about nudging. For the remainder of this thesis, I will use the term “nudge” to refer to this general category of interventions based on insights from behavioral sciences. When needed, examples of specific interventions will be given.

On a policy level, nudge interventions typically benefit from being cheap and easy to implement, resulting in a cost-effective way of inducing behavior change. This has made nudging a popular tool for businesses and governments. A recent report by the OECD documents that nudges or similarly behaviorally informed interventions are used by over 200 governmental units and initiatives around the world (OECD, 2019). Several countries have even created specialized governmental “nudge units,” groups of experts working to support other parts of government in applying behavioral insights to societal problems (for a history of the development of one prominent unit, see Halpern, 2015). In the best-case scenario, nudging can be an effective and freedom-preserving alternative for policy. As we will next consider, however, critics dispute the success in achieving both these aims.

Challenges for nudging

Ideally, policy interventions will be successful and effective in steering behavior while at the same time restraining freedom of choice as little as possible. As noted however, there can often be a negative connection between the two goals, with soft interventions preserving freedom but lacking effectiveness, and hard interventions being efficient but intrusive. While nudging has been championed as a potential golden middle way (Thaler & Sunstein, 2008), the framework has nevertheless ended up fighting a two-front battle, criticized on both grounds.

In his book detailing the development of the United Kingdom’s Behavioural Insights Team, chief executive David Halpern (2015, p. 304) noted that the political right criticizes the nudge approach for being “too close to the dark arts of propaganda and subconscious manipulation,” while the left fears that nudging is “an excuse for not acting more decisively and effectively.” Harsher still, philosopher Sarah Conly (2013) has portrayed nudging as failing both to respect individual autonomy and to be sufficiently effective in influencing behavior. Conly noted that, compared with traditional soft and hard

policy measures, “libertarian paternalism” (Sunstein & Thaler, 2003) and nudging may be “in a sense, the worst of both worlds” (Conly, 2013, p. 8).

Effectiveness issues

Whether nudges ought to be considered effective depends on what the relevant frame of reference is. Nudging may, for instance, be evaluated in relation to the effectiveness of other policy measures, in relation to its cost-effectiveness, or in relation to the effect sizes commonly found in similar psychological research. Depending on the perspective, the verdict on nudging can differ dramatically.

Critics skeptical of the effectiveness of nudging seem to focus on the first comparison, i.e., comparisons with other policy measures, and in the context of tackling major societal challenges. For instance, Goodwin (2012, pp. 89–90) expressed the fear that “that nudging is unable to deliver the kind of substantive changes that are needed to tackle the big problems that society faces.” To the best of my knowledge, no systematic or large-scale comparisons between the absolute effectiveness levels of nudge interventions and other policy interventions exist. A review from 2015 claimed that none existed at that point in time (Kosters & Van der Heijden, 2015, p. 285), and I have been unable to find any relevant work published later. Without this knowledge, the charge is impossible to support or disprove empirically. Speculatively, it seems plausible that nudge interventions will be less effective in changing behaviors than harder measures such as taxes, incentives, or bans. At the same time, common nudges likely match or outperform other soft interventions such as information campaigns (which some also consider nudging, see Sunstein, 2018). Whether this validates or refutes concerns about comparative effectiveness thus depends on the level of expectations.

A second approach to evaluating effectiveness is in relation to the costs of implementation. As many nudges are relatively cheap to put into practice, applying a cost–benefit perspective produces a more favorable outcome for nudging. A study by Benartzi et al. (2017) analyzed data on interventions in energy conservation, health, retirement savings, and more. In all instances in which the outcome was assessed in monetary amounts, the relative effectiveness of the nudge was positive. Interestingly, the study also made comparisons with the cost-effectiveness of other types of policies, and found that nudge interventions typically excelled. For instance, in relation to energy conservation, a social norm intervention in which households were nudged by being supplied information on neighbors’ electricity usage (Allcott, 2011) saved 27.3 kWh per dollar spent. In comparison, one traditional incentive-

policy of providing discounts on electricity for households that managed to cut their summer electricity usage by 20% (Ito, 2015) saved 3.41 kWh per dollar spent. Another incentive program to reduce energy consumption during peak times (Arimura, Li, Newell, & Palmer, 2012) saved 14.0 kWh per dollar spent, which is still only about half the relative effectiveness of the social norm nudge. By adopting a cost–benefit perspective, nudging appears highly defensible as a policy tool with regard to effectiveness.

A third approach is evaluating the effects of nudge interventions in relation to comparable phenomena within behavioral science. For instance, a recent study compiling the results of 134 meta-analyses of phenomena in social psychology concluded that effect sizes at the 25th, 50th, and 75th percentiles corresponded to Cohen’s d s of 0.15, 0.36, and 0.65 (Lovakov & Agadullina, 2021). In this comparison, common nudge interventions performs well. A meta-analysis of default interventions by Jachimowicz et al. (2019) found an average effect of $d = 0.68$. For social norm interventions, a meta-analysis of field experiments on pro-environmental behavior found an average effect of $d = 0.32$ (Bergquist, Nilsson, & Schultz, 2019). Again, however, it should be noted that there is considerable heterogeneity between the effect sizes of different nudge interventions, and that default interventions on average clearly outperform other intervention types (Hummel & Maedche, 2019). A word of caution is also warranted, in that some recent research has questioned the level of scalability for nudges from lab experiments to large-scale public policies (Al-Ubaydli, Lee, List, Mackevicius, & Suskind, 2021). Comparing results of governmental field trials with results of research studies, DellaVigna and Linos (2020) found an average effect of about one quarter of what had been expected.

In conclusion, the extent to which nudging stands up to traditional policy tools in terms of absolute effectiveness is hard to determine. Concerns may be valid if nudging is cast as a substitute for harder interventions, or expected to mitigate major societal problems on its own. However, the former appears avoidable and the latter perhaps an unreasonably big requirement. When assessed in relation to cost-effectiveness, or to what can be expected regarding behavioral science phenomena, criticism based on alleged ineffectiveness seems overblown.

Ethical issues

Although most of the behavior change interventions championed in *Nudge* (Thaler & Sunstein, 2008) have been around for a long time, the book’s publication led to an explosion in thinking and writing about the legitimacy of behavior change interventions based on insights into human psychology. The

last decade has seen the publication of numerous academic papers (see Delaney, 2020, for a list of over a hundred journal articles and book chapters on the ethical aspects of behavioral science and policy) and several books that dissect the topic in detail (e.g., Conly, 2013; John, 2018; Rebonato, 2012; Sunstein, 2014, 2015, 2016; Sunstein & Reisch, 2019; White, 2013; see also anthologies edited by Coons and Weber on manipulation, 2014, and paternalism, 2013, which contain many direct and indirect links). In addition, several frameworks for the ethical application of nudges have been developed (Clavien, 2018; Lades & Delaney, 2019; OECD, 2019; Sunstein & Reisch, 2019). This section will sketch the central lines of criticism that nudging has evoked, and some responses to these criticisms. The aim is not to assess the normative force of the theoretical arguments on either side, but to provide background on issues that will later be investigated empirically. For a recent overview and assessments of relevant ethical worries, see Schmidt and Engelen (2020).

What is the problem?

The central ethical concern with nudging concerns autonomy (Schmidt & Engelen, 2020; Sunstein, 2016). Critics see nudging as unduly interfering with people's sovereignty over their lives and decisions. These worries cluster to form two main lines. One line focuses on the ends to which a person subjected to a nudge is steered (or whether a person ought to be steered toward any ends at all). This can be summarized as the concern that nudging is paternalistic. A second line focuses on the means by which nudging influences a person, that is, the psychological processes and mechanisms that common nudges use to induce behavior change. The concern here is that the means by which nudges work constitute a "perversion of the decision-making process" (Wilkinson, 2013, p. 345), or in other words, that nudging is manipulative.

In both cases, critics claim, the nudger is insufficiently respecting the nudgee's capacity to choose for themselves.

Paternalism

Worries about paternalism may arise in situations in which one person is deciding for another. In brief, paternalism means that an agent influences another agent's choices without the latter's explicit consent, in a direction deemed desirable by the first agent (for a discussion of fleshed-out definitions, see Dworkin, 2013). Even if the influence results in an outcome that is beneficial to the affected, an anti-paternalist will argue that the influencer's behavior is problematic by virtue of their deciding *for* the other, since that

simply is “none of their business” (Coons & Weber, 2013, p. 5). It is thus not hard to see how common nudges are troublesome in this regard. In the typical nudge scenario, a choice architect will have decided what course of action is to be encouraged by a nudge, and the intervention will attempt to influence the nudgee in this direction regardless of their own preference in the matter. For instance, if an employer institutes automatic enrollment in a retirement savings plan for new employees (Madrian & Shea, 2001), this implies that the employer considers saving a certain amount for retirement to be a goal that the employee would be better off having (and may fail to achieve without the nudge). Consequently, the actions of the employer and nudger can be deemed problematic, as their attempt to decide for the employee may be taken as a lack of respect for the employee’s capacity to decide for themselves—in other words, their autonomy.

Manipulation

Manipulation, on the other hand, principally has to do with undue influence on the process by which a decision is made (e.g., Sunstein, 2016; Wilkinson, 2013). Wilkinson (2013, p. 345) noted that “whereas coercion uses threats, which involve changing the costs of selecting certain options, manipulation involves some underhand interference with the ways in which people see their options,” and characterized manipulation as a “perversion of the decision-making process.” In other words, manipulation is a threat to autonomy because it leads people to choose differently than they otherwise would have, and for reasons they may be unaware of. For instance, people are often highly susceptible to what other people do or approve of (i.e., social norms) but fail to realize the extent to which such social information affects their behavior (Nolan, Schultz, Cialdini, Goldstein, & Griskevicius, 2008; Jaeger & Schultz, 2017).

Responses to criticisms

Several responses to criticism were already offered in *Nudge*. In it, Thaler and Sunstein (2008) argued that due to the malleability of the human mind, some level of environmental influence is inescapable (“there is no such thing as a ‘neutral’ design,” p. 3). Furthermore, they argued, benevolent intervention is preferable to random design. Notably, neither of these arguments attempts to refute claims of manipulation or paternalism head on, but settles for arguing that we should not be overly concerned if some interference takes place (referring to the fact that some interference “must” happen inevitably). Another response given is that nudging should only be used in a way that leaves nudgees

“better off, as judged by themselves” (Thaler & Sunstein, 2008, p. 5). In this way, the goals of the nudgee would never be overruled by nudging (although a person’s means of achieving their goals might be).

In relation to manipulation, Thaler and Sunstein (2008) further stressed that nudges should always be easy and cheap to avoid, so as not to coerce behavior. Whether or not this succeeds may be a question from case to case (Wilkinson, 2013). Another line of response is that the use of nudges should always be transparent to the people subjected to them. How openly and “directly” transparent information ought to be in order to avoid manipulation has been debated (Bovens, 2009; Ivankovic & Engelen, 2019). One account claims that it may suffice to provide a general disclosure in public records that certain nudges are in use and to what end (cf. discussion in Thaler & Sunstein, 2008, pp. 246–249). A stronger account may require that the nudge is detectable at the time it is exerting influence, as judged either by an average person or by an attentive, “watchful” person (Bovens, 2009, and Ivankovic & Engelen, 2019, both argue in favor of the latter). Interestingly, while some have considered the latter approach extreme (Hansen & Jespersen, 2013), it has been the main route taken by research –including in this thesis– that has empirically investigated transparency when subjecting people to nudges.

People’s perceptions and experiences of nudges as partial solutions: A framework

The previous section identified both ethical and effectiveness concerns as central challenges for nudging as a valid policy measure. In this section, I propose that people’s perceptions and experiences of nudges are underappreciated features of the intervention that can provide partial solutions to both issues. In the framework presented here, I suggest that people’s perceptions, and especially first-hand experiences, of nudges are important in at least three ways: (1) to inform debates about the ethics of nudging; (2) as part of the net effect of the intervention; and (3) as a moderator of the nudge’s success in influencing behavior. In other words, by properly appreciating what people think of nudges, and experience as a consequence of the nudge, researchers and policy makers will be better equipped to assess the validity and value of the intervention. The empirical studies of this thesis closely connect to the framework, as will be described on pages 28–29.

Separate parts of this framework have been raised earlier by other authors. However, to my knowledge, the framework is novel in its aim to provide a broad account of how people’s perceptions and experiences of nudges are

informative for policymaking. The framework will in turn function as a lodestar for the subsequent literature review, and provides a rationale for my own empirical work presented later in this thesis.

In addition to listing three main ways in which perceptions and experiences are valuable to evaluation of nudges, this section will highlight the role of nudge transparency. Nudge transparency can affect how people experience and perceive nudges; if a nudge is insufficiently transparent to be noticed, no perceptions or experiences of the nudge will ensue. Efforts to make nudges transparent, such as by disclosing the nudge, may furthermore be construed as part of the intervention and have corresponding consequences for the three principles in the main framework. However, since nudge transparency is not a way in which a nudge experience or perception is valuable, I do not include it as a principle in the framework.

Principle 1: People's perceptions and experiences inform the ethical assessment of nudges

Apart from in work by Sunstein (e.g., Sunstein, 2016; Sunstein & Reisch, 2019), the nudgee's perspective has largely been missing in debates on the ethics of nudging. I argue there are at least two senses in which people's perceptions and experiences can inform ethical discussions. First, nudgees' opinions of what and when interventions are ethical, fair, and acceptable can be considered a relevant angle in ethical discussions. This may be referred to as the policy acceptance-sense, and has been the dominant focus in empirical research so far. Opinion data of this kind has been held to be a central guiding principle for the ethical application of nudges by several sources, for instance, in governmental (House of Lords, 2011), organizational (OECD, 2019), and academic frameworks (Lades & Delaney, 2019; Sunstein & Reisch, 2019). Second, data can be used to assess the empirical feasibility of premises or conclusions in normative arguments. For instance, if a nudge is thought to be unethical because it infringes on the autonomy of decision makers, or makes them less happy, providing empirical evidence that decision makers experience lack of decisional control or dissatisfaction would strengthen the argument. In contrast, if repeatedly and under circumstances of good measurement, the nudge is not found to reduce decision makers' sense of control or satisfaction, the same argument loses credibility.

That either of the two sense described above has much value may of course be denied by a skeptic. Someone who might be labeled an "armchair traditionalist" (Kauppinen, 2014) as to the normative discussion would likely object by claiming that, in relation to ethics, expert theoretical analysis trumps

anything that can be learned by empirical research. For instance, because the ethical analysis of the general public is feared to be shallow or biased (but see Horvath & Wiegmann, 2021, for evidence that expert ethicists can be susceptible to the same judgment biases as can laypeople). On this position, assessing public opinion may nevertheless be pragmatically valuable to ensure the perceived legitimacy of policies.

However, at least among proponents of nudging, it appears common to believe that people subjected to nudges themselves have a privileged standing when it comes to assessing the nudge and its outcomes. This is captured in the oft-used phrase that people should be “better off, *as judged by themselves*” (Thaler & Sunstein, 2008, p. 5, my emphasis). Applied to ethical matters, this sentiment entails adherence to ethical empiricism (Kauppinen, 2014), a position holding that the proper way to settle ethical issues is not solely on theoretical grounds; instead, matters will to some extent be empirical. Apart from an adherence to empirics, the “better off, as judged by themselves” perspective also entails what can be labeled a subjectivist position¹ as to who should be the judge of a person’s own fortune. According to a strong subjectivist account, the nudgee’s perception of their situation should be considered the decisive criterion for assessing, for instance, whether a nudge infringes on their autonomy. In other words, the ethicality of a nudge would greatly or completely depend on the view of the person subject to it. Thereby, research on how people experience nudges would be a fundamental interjection into the ethical discussion (“bold” ethical empiricism, in Kauppinen’s terminology). According to a more modest account, empirical data can still be considered a useful indicator of ethical matters, although not the sole or finally decisive one.

In sum, if some level of ethical empiricism is accepted, then people’s perspective on how nudges affect their decision making ought to be acknowledged in assessing the ethics of nudging. If ethical empiricism is rejected, research on how people perceive and experience nudges is still informative for legitimacy. As will be discussed below (pp. 24–25), I argue that there is an important distinction to be made between people’s opinions of nudges and people’s experiences of nudges, with the former being indirect and therefore insufficient in illuminating ethical issues in the second sense discussed here.

¹ I want to stress that in the section of *Nudge* quoted above, Thaler and Sunstein (2008) advocate the “as judged by themselves” criterion, i.e., an empirical and subjectivist position, in relation to preferences for goods (“apples or pears”), and not necessarily for ethical assessment.

Principle 2: People's perceptions and experiences are part of the net effect of the nudge

There is more to the success of a nudge intervention than whether it influences the targeted behavior. Changes in perceptions and mental states of nudgees brought on by the intervention, whether positive or negative, ought to be counted toward the net effect of a nudge (for a similar argument, see Allcott & Kessler, 2019; Thunström, 2019). For example, many nudges attempt to make the target behavior more fun or easy to engage in (Behavioural Insights Team, 2014). If a nudge produces the same behavior as another intervention (or no intervention) but makes the decision more pleasant, the net effect of the nudge is higher. For example, consider the nudge to reduce littering in which smokers are given the choice of putting out their cigarette butts in one of two containers, the choice of container constituting a vote, for example, between rival sports teams (see <https://ballotbin.co.uk/>). Another sense is whether a nudge renders a hard choice easier to make, for example, by making information easier to process (Larrick & Soll, 2008) or by providing an implicit choice recommendation (McKenzie et al., 2006). In this case, time and effort are saved for the decision maker, and frustration and regret are potentially avoided. When a nudge is perceived as helpful, the person subjected to it may furthermore think more highly of the nudger. If the choice architect is perceived favorably as a result of implementing the nudge, this is also a positive addition to the net effect.

While certain nudges and certain circumstances may be well situated to produce beneficial outcomes in the form of positive perceptions and experiences, the opposite can also be the case. Due to fears that choice experiences can be degraded by a nudge, it may often be as important to investigate whether a nudge intervention produces negative perceptions or experiences. For instance, a nudge may produce irritation, anger, or disappointment if the nudgee construes it as an illicit influence attempt. The feelings of people with strong “Hands off!” attitudes toward outside interference may thus risk reducing the net effect of a nudge. Another case is when a nudge may shame people into engaging in a behavior, for example, by invoking a norm that the nudgee feels obliged to uphold.

Even though a few exceptions exist (Allcott & Kessler, 2019), the mental states of people subjected to nudges have typically not been empirically assessed and accounted for in summaries of the net effect of nudge interventions (e.g., Benartzi et al., 2017; Cadario & Chandon, 2020; Hummel & Maedsche, 2019). Evaluations of the net effect of nudges therefore risk overestimating the value of interventions that are successful in influencing

behavior but have negative side effects in terms of the perceptions and experiences of those affected (cf. Loewenstein & O'Donoghue, 2006). Correspondingly, the worth of nudges that facilitate decision-making or render choosing more enjoyable may commonly be undervalued.

Principle 3: People's perceptions and experiences can moderate behavior change

An understanding of how people perceive and experience a nudge intervention is furthermore relevant to behavioral outcomes (a similar argument has been made by Krijnen, Tannenbaum, & Fox, 2017). In a direct sense, whether or not people find a nudge acceptable will likely influence its implementation success; if people disagree with an intervention, they are likely to seek to avoid it. As noted by Hagman (2018, p. 34), "if the general public would find nudges to be unacceptable, then it would not matter whether nudges are theoretically intrusive to freedom of choice or not." Indirectly, how a nudge is perceived will also affect the social interplay between the nudgee and the nudger (Krijnen et al., 2017). Apart from the potential direct benefit as a secondary outcome noted above, this may in turn affect the success of subsequent interventions through a reciprocity mechanism. For instance, even if a nudge is successful in influencing an initial behavior, if its use leads to negative perceptions of the nudger, an intervention by the same nudger at a later stage may be less influential or even be rejected outright. Conversely, if the nudge is designed and used in a way that is perceived positively by the nudgee, this could increase trust in the nudger and affect subsequent interactions positively.

The relevance of intervention transparency

As discussed in previous sections, nudges typically make greater use of refined psychological insights to drive behavior change than traditional policy interventions do. Because of this, their influence can be harder for subjected individuals to realize and sometimes the intervention may not be noticed at all (for instance, default nudges can be easy to miss). For this reason, a nudge's level of transparency, both whether it is noticed and its influence mechanisms understood, is closely related to how it will be perceived and experienced, and below I mention two senses. As will be seen in the literature review next, transparency has also been an important theme in the empirical research on perceptions and experiences of nudges, with some strains hard to separate from the researchers' active manipulation of transparency.

Transparency as an indirect catalyst of perceptions and experiences

In any real-world application, a nudge's impact on the three principles described above is likely moderated by its level of transparency. As many common nudge interventions are considered largely non-transparent (e.g., Hansen & Jespersen, 2013), not the least default nudges (Smith, Goldstein, & Johnson, 2013), the risk of the intervention being unnoticed is significant. For instance, showing that people subjected to a nudge experience their autonomy as intact will convince few skeptics if the nudges were unaware of the nudge's presence and potential influence. Similarly, potentially beneficial secondary outcomes or downstream consequences may be less likely to occur if the nudge passes unnoticed. This may be a contributing reason why research on the spillover effects of nudge interventions have so far found only small effects (Hedesström, Michaelsen, Nyström, Luke, & Johansson, 2019).

Transparency in its own right

Of course, under-detectability is also at the core of the worry about manipulation discussed previously. In this way, ensuring that a nudge is made transparent to the person affected by it is a way of alleviating concerns of manipulation, thereby making the intervention more ethical.

Making a nudge transparent may also on its own, rather than the characteristics of the nudge itself, be a driver of positive perceptions of the nudge and nudger. Actively disclosing a nudge may, for instance, signal trustworthiness (cf. Paunov, Wänke, & Vogel, 2019a), and evoke a willingness to interact with the choice architect again (Steffel, Williams, & Pogacar, 2016).

How do people perceive and experience nudges?

Initially, empirical research on nudges had a strong focus on the effectiveness of the nudge in influencing the target behavior. Recent years have, however, seen increasing interest in how people perceive and experience nudges. This latter line of research has taken one of two approaches. One approach has been to ask participants to judge descriptions of nudges presented to them in survey studies. Here, people's opinions are based on detailed, prototypical cases of nudges, but ratings are given without participants actually encountering the

nudge. In the remainder of this thesis, this approach will be referred to as a *description-based approach* to researching perceptions of nudges. A second approach has been to subject participants to nudges first hand. In other words, if the intervention of interest is a default nudge, participants have been asked to make a decision in a choice situation containing a default nudge, and afterwards asked for their perceptions or experiences. Here people lack the birds-eye view of the description-based approach, but their perceptions and experiences reflect real encounters with nudges as actual decision-making is taking place. In the remainder, this approach will be referred to as an *experience-based approach*.

Description-based research

The description-based research approach has so far been the dominant paradigm of research into people's views of nudges. In these studies, participants are typically presented with a concrete nudge application that details the purpose of the nudge and the mechanisms by which it operates. For example, the following is the default nudge-organ donation vignette shown to participants in a study by Hagman, Andersson, Västfjäll, and Tinghög (2015, p. 444):

There is currently a lack of organ donors in many countries. In some places, to become an organ donor the individual has to make an active choice and register as an organ donor with the appropriate authority. If no choice is registered, the individual is assumed to be unwilling to donate in event of an accident (so-called Opt-In). In previous surveys most people report that they are willing to be an organ donor but have not registered.

One way to increase the number of organ donors could be to automatically enroll people as organ donors unless otherwise specified (so-called Opt-Out). In other words, it is up to the individual to register at the appropriate authority if they are unwilling to donate their organs in the event of an accident. The aim with this intervention (Opt-Out) is to increase the number of organ donors.

After reading about the nudge, participants are asked to rate the intervention on one or more dimensions. Of primary interest in most research has been the acceptability of the nudge. However, several studies have also explored further facets of acceptability, such as people's worries and concerns, for instance, the

extent to which people find various nudges intrusive to freedom of choice, or whether they project that the nudge would lead to less authentic decision-making.

The overall finding is that most researched nudge interventions receive majority support in most studied countries (Hagman et al., 2015; Jung & Mellers, 2016; Pe'er et al., 2019; Reisch & Sunstein, 2016; Sunstein, Reisch, & Rauber, 2017). However, variation in acceptability ratings has been found depending on: the country studied² (Hagman et al., 2015; Reisch & Sunstein, 2016; Sunstein et al., 2017); type of nudge and application context (e.g., Jung & Mellers, 2016; Sunstein et al., 2017); whether alternatives to the nudge are given (Davidai & Shafir, 2020; Hagman, Erlandsson, Dickert, Tinghög, & Västfjäll, 2019); personality characteristics of the rater (Jung & Mellers, 2016; Hagman et al., 2015; Nilsson, Erlandsson, Västfjäll, & Tinghög, 2020); and who the nudge designer is (Gold, Lin, Ashcroft, & Osman, 2020; Tannenbaum, Fox, & Rogers, 2017). In addition, and relevant to this thesis, intervention transparency has been found to be one of the strongest predictors of acceptability.

A handful of studies has compared nudges that are transparent and target reflective thinking with nudges that are non-transparent and target subconscious mental processes. Across countries and application contexts, the highest support levels are found for transparent nudges that work by information provision (e.g., an education campaign) or by making existing information more salient or easier to grasp (e.g., visualizing the healthfulness of foods by coloring them akin to a traffic light) (Reisch & Sunstein, 2016; Sunstein et al., 2017). Such informational nudges have typically received approval ratings of around 75–95%. The same studies found that default rules (e.g., a default to donate a portion of one's tax return to charity) also received majority support, but with lower approval overall and a higher dispersion of ratings of 30–90%.

Other studies have researched how transparent and non-transparent nudges compare when aiming to affect one and the same behavior. For instance, Gold et al. (2020) looked at applications in exercise, eating behavior, alcohol consumption, smoking, and financial choices, and found that participants rated

² With few exceptions (Pe'er et al., 2019; Sunstein et al., 2017), the research reviewed in this section has used samples from the EU, UK, and the US. It should be kept in mind that the use of samples from predominantly western individualistic societies may limit the generalizability of the findings. For instance, Sunstein et al. (2017) noted that China and South Korea stand out in their results by showing “spectacularly high approval rates” (p. 1) compared with western countries, evidencing heterogeneity between cultures.

the transparent interventions more acceptable in all instances.³ The difference was especially large when the interventions targeted monetary behaviors. Here, mean acceptability between the transparent and non-transparent interventions differed by approximately three points on a nine-point scale. Jung and Mellers (2016) obtained similar results: comparing transparent and non-transparent nudges applied in medical plan enrollment, retirement savings, credit card reimbursement, organ donation, and water conservation, a consistent preference for the transparent nudges was found.

In addition to acceptability, a few studies have researched perceptions of ethical dimensions of nudges. Jung and Mellers (2016) found concerns about threat to autonomy and paternalism regarding both transparent and non-transparent nudges, but that the concerns were highest for the non-transparent ones. Relatedly, Felsen, Castelo, and Reiner (2013) found that the more the nudge was perceived to target reflective thinking, the more participants predicted that choices would feel authentic.

The less favorable perceptions of non-transparent interventions hint at a potential issue for efficient nudging, as the most effective nudges, default nudges, are often considered to fall in this category (Hansen & Jespersen, 2013; Smith, Goldstein, & Johnson, 2013). Indeed, the two default interventions researched by Hagman et al. (2015)—default enlisting of organ donors and a default to compensate for the carbon emissions of a flight—were the only ones to just barely receive majority support, and more than three quarters of the sample found them intrusive to freedom of choice. Yan and Yates (2019) investigated whether default nudges could be perceived more favorably if the choice architect made efforts to make the nudges more ethical. However, despite ethically intended measures such as disclosing the intention behind the intervention, or emphasizing how to opt out from it, they found that the default nudges were perceived as more manipulative and were anticipated to harm autonomy more than the contrasting opt-in policy. Whether the enhancement improved perceptions of the default nudge was not clear, however, as no “non-enhanced” default control condition was included in the study.

In sum, people are generally positive toward nudges but have concerns about interventions when they are less transparent and when they incur losses. Although most studies do not allow for a full disentanglement of the type of nudge and its application context, default nudges are typically rated least favorably. This is unfortunate from a policy perspective, as defaults tend to be more influential than other nudges in changing behavior (Hummel & Maedche, 2019). It is discouraging that efforts to make default nudges more ethical appear unsuccessful (Yan & Yates, 2019), although it remains unclear whether

³ However, for exercise the difference failed to reach significance after the authors' Bonferroni correction.

this should be interpreted as people simply preferring opt-in defaults, or because the investigated enhancements lacked an effect. As we will see next, the literature on how people perceive nudges when subjected to the interventions first hand have predominantly focused on defaults, and the findings of this approach appear somewhat more favorable from the perspective of nudging.

Experience-based research

Somewhat surprisingly, not much has been published on the subjective experiences of being nudged. This may be due to the relative youth of the nudging framework, and that research so far primarily has been occupied with establishing the effectiveness of interventions in modifying target behavior. Another reason may be that there has been a tendency to view nudging as a largely covert category of operations (e.g., Bovens, 2009). If it is thought that nudge interventions are not detected by nudgees, it would make a lot less sense to ask people how they felt after having “experienced” them. A recent stream of experimental studies has, however, started to investigate the issue of transparency in nudge interventions, and it is primarily from these studies that data on experiences of being subjected to nudges are available.

The following section is divided into two parts. The first deals with what I call *choice experiences*, which have to do with the experiences and feelings of a nudgee in relation to being subjected to the nudge. The second part deals with perceptions of the nudge intervention and the choice architect (i.e., the person implementing the nudge).

Choice experiences

Choice experiences are central to all of the principles described in the framework above, and can only be properly investigated using an experience-based approach. For instance, when using an experience-based approach one can directly measure how autonomous or satisfied an individual feels when making a choice subjected to a nudge. This is in contrast to the description-based approach that is confined to indirect measurement regarding experiences and mental states; limited to estimations and guesses.

For the ethically central issue of how nudge interventions affect people’s experienced level of autonomy, choice experience data exist from three previous studies. Two of these addressed the issue head on by comparing experienced autonomy for people subjected to a nudge versus making the same choice without the nudge. Unfortunately, however, some limitations exist as to

what conclusions can be drawn from these studies, which will be discussed below. The third study explored how experienced autonomy is affected by making the nudge more transparent (vs. not transparent). All three studies used opt-out defaults to nudge participants.

In one of the studies of the first kind, Arvanitis, Kalliris, and Kaminiotis (2020) found suggestive evidence that default nudges may be detrimental to people's experienced autonomy. In the experimental task, participants were faced with a hypothetical choice of health insurance plans. The researchers manipulated how many insurance plans there were to choose from, three or nine, and whether or not one of them was made the default. Results showed that when there were three health plans to choose from, participants subjected to the default nudge gave significantly lower ratings for one subset of the autonomy scale (the "locus of causality" subscale). The magnitude of the effect was notably strong, and corresponded to a Cohen's d of approximately 0.95 (near the 90th percentile of effect size magnitudes in social psychology research, Lovakov & Agadullina, 2021). This finding seemingly bolsters ethical criticisms. However, I believe there are reasons for caution in drawing such conclusions based on these data. First, no differences were reported for the autonomy scale as a whole, or for either of the other two subscales it encompassed (i.e., the "perceived choice" and "volition" scales). This lack of convergent results are unexpected given the magnitude of the effect on the locus of causality (when there was three choice options). Second, virtually no difference was found between participants subjected to and not subjected to the nudge when nine choice options were presented. Together, this raises concerns about the robustness of the finding that a default nudge diminishes people's experienced autonomy. Given the limited sample size of the study (total $N = 139$, or approximately 35 per cell), I suggest that a preregistered replication in a larger sample is needed before ethical and policy conclusions should be speculated on.

In the second study, Abhyankar, Summers, Velikova, and Bekker (2014) provided suggestive evidence that a default nudge's influence on experienced autonomy may not be negative after all. In a hypothetical experiment on cancer treatment choices, participants were asked about the perceived behavioral control over their choice after having been subjected to a default manipulation. Specifically, participants chose whether to take part in a clinical trial treatment, and had the question framed in an opt-in, opt-out, or active choice format (i.e., no preselected option). The results showed high average perceptions of control for participants subjected to the opt-out default nudge, and no differences between choice formats. Participants were also asked how satisfied they were with their choice, with similar results. However, from a nudge influence perspective, a limitation exists in that the study did not focus solely on the

choice format (“nudge”) influence, but sought to evaluate a broader choice process. This choice process also included a second not-defaulted opportunity for participants to state their preference, which is why it is not possible to isolate the effect of the nudge. Thus, it is not clear whether or not the default nudge actually influenced autonomy or satisfaction, as any effects may have been mitigated by the subsequent decision process.

Third, a study by Wachner, Adriaanse, and De Ridder (2020) investigated whether making a default nudge more transparent to the decision maker affected their experienced autonomy. The authors reasoned that even though previous studies on disclosing nudges (see next section) have detected little influence on people’s choices, autonomy may still be affected, especially when the nudge disclosure highlights opaqueness in the nudge. In the experiment, participants were asked to make a choice of whether to participate in an extended version of the study for no additional pay. For all participants, the default was set to accepting the long questionnaire version, but level of transparency of the nudge was manipulated in three different ways. One experimental condition presented a disclosure of the nudge, which read “Please note the preselected default option. It is meant to encourage people to choose the longer version of this questionnaire” (p. 5). A second condition presented the same disclosure, but with the following sentence added: “People are usually unaware of its influence” (p. 5). A third, control, condition did not present any explicit information on the presence of the nudge. It was hypothesized that participants receiving the short disclosure (not highlighting opaqueness of the nudge) would score highest on autonomy, participants without a disclosure second, and that participants receiving the long disclosure highlighting opaqueness would score lowest. The results, however, revealed no significant differences between the three conditions, and the same was found for a measure of participants’ satisfaction with their choice. As in the study by Abhyankar et al. (2014), reports of autonomy and satisfaction were high, which speaks against a negative influence of nudging on people’s choice experiences.

Apart from autonomy, a few studies have investigated choice experiences in the form of the emotive consequences of being subjected to nudges. Bruns, Kantorowicz-Reznichenko, Klement, Luistro Jonsson, and Rahali (2018) subjected participants to a default nudge in a climate compensation scenario. Even though the default nudged participants to donate part of their study reimbursement, it did not lead to feelings of irritation, anger, annoyance, or aggravation. However, there was no control group for the emotion measures, leaving it unclear whether the default nudge may nonetheless have influenced participants’ feelings.

Hedlin and Sunstein (2016), in a hypothetical scenario concerning household energy provision, found that nudging participants toward green energy by making this option the default led to lower levels of guilt at not enrolling than when the choice had no default. However, the difference was small in magnitude and reported levels of guilt were low in both conditions, just crossing the neutral midpoint of the scale.

Lastly, some nudges use the induction of negative emotions as an intentional and integral part of the intervention. For instance, Thunström (2019) studied a salience nudge in which calorie labeling was highlighted for food choices, and found that participants low in eating self-control were emotionally taxed by the nudge. Other examples include graphic warnings, such as on cigarette packaging, and social norm interventions, which can shame people into certain behaviors (Thøgersen, 2006). I will not further review intervention studies based on emotional taxation here, but note that when nudges work by this mechanism, it ought to be considered part of the net effect of the intervention (cf. argumentation in Thunström, 2019) and might be less likely to foster a positive relationship between the choice architect and nudgee.

Perceptions of the nudge and the choice architect

Furthermore, several studies are not directly concerned with experiences of choosing but rather with people's perceptions of the nudge and of nudger. For instance, Bruns and colleagues (2018; Bruns & Perino, 2019) investigated the perceived threat to freedom of choice from a default nudge. In two experiments on carbon offset contributions, they found that a default nudge was perceived as more freedom threatening than a recommendation (but less so than a mandate, Bruns & Perino, 2019), and that increasing the transparency of the nudge did not affect how it was perceived (Bruns et al., 2018). Unfortunately, neither study included a non-nudged control condition for the perceived threat to freedom of choice measure. In relation to ethical concerns, it should be noted that while freedom of choice is a prerequisite for autonomy, perceived threat to freedom of choice is a judgment of whether or not an influence attempt is happening (at least as defined in the scale used here; see also Study I of the present thesis). Consequently, this is different from the influence actually having been exerted on a choice, which is what would constitute an infringement on autonomy. Relatedly, the aforementioned study by Wachner et al. (2020) measured whether the nudge was perceived by participants as trying to pressure them into a choice, but no differences were found between the disclosed and non-disclosed conditions.

Steffel, Williams, and Pogacar (2016) found, using a hypothetical apartment acquisition scenario, that participants perceived being nudged toward pro-environmental amenity choices to be fair (vs. unfair). They also found that participants were more willing to work with the choice architect again in the future if the choice architect had disclosed the choice format to them.

Other examples of whether a default nudge intervention, or rather a transparent versus non-transparent default nudge, can affect perceptions of the choice architect were found by Paunov, Wänke, and Vogel. In two experiments with hypothetical choices of elective university courses, mixed evidence was found in that increased transparency did not influence the perceived trustworthiness of the nudger (Study 1), but did increase perceptions of being deceived (Study 2; Paunov, Wänke, & Vogel, 2019a). However, a second study with a defaulted choice of survey length (cf. Wachner et al., 2020, above) found no differences in perceived deceptiveness (Paunov, Wänke, & Vogel, 2019b).

In sum, people seem to perceive nudges differently depending on the evaluation mode in which their judgments are elicited. In experience-based studies, findings are arguably somewhat more favorable than in description-based studies. This has important consequences for how people's opinions of nudges should be assessed, as one and the same nudge may be considered acceptable when judged in one mode, but unacceptable when judged in another. Thus, policy decisions guided by only a single perspective risk misrepresenting public perceptions.

Transparency and choice

Studies conducted within the experience-based paradigm have additionally documented how increasing intervention transparency affects the nudge's influence on choice. If the nudge is still effective in changing behavior when made transparent, it may be feared that manipulation has not been mitigated (Steffel et al., 2016). However, this is not necessarily the case. The nudgee may still perceive the nudge as a legitimate course of action, and a difference in behavior between a nudged and non-nudged individual could reflect that the nudge merely succeeded in facilitating the behavior (Ivanković & Engelen, 2019).

As for the studies reviewed above, research into transparency effects on choice consist of studies in which transparency has been manipulated by means of a nudge disclosure. The near-universal finding is that when a disclosure is presented along with the nudge, this does not diminish the effectiveness of the

intervention in regard to changing behavior. So far this has been tested in hypothetical choice scenarios concerning end-of-life treatments (Loewenstein, Bryce, Hagmann, & Rajpal, 2015), sharing personal data to help improve health care (Dranseika & Piasecki, 2020), sharing personal data on social media (Steffel et al., 2016), choice of environmentally friendly or premium apartment amenities (Steffel et al., 2016), and choice of elective university courses (Paunov et al., 2019a). In consequential choices in “lab settings,” transparency effects on choice have been tested in contexts concerning charity donations for carbon offsets (Bruns et al., 2018), participation in additional research surveys (Steffel et al., 2016), and participation in longer research surveys (Paunov et al., 2019a, 2019b, 2020; Wachner et al., 2020). Lastly, a few studies have employed consequential choices in field settings concerning healthy food choices (Cheung, Gillebaart, Kroese, Marchiori, Fennis, & De Ridder, 2019; Kroese, Marchiori, & De Ridder, 2015; Steffel et al., 2016).

Notable exceptions to the otherwise dominant null effects are the studies by Paunov et al. (2019a; 2019b; 2020), which found an increased effectiveness for disclosed nudges. A second thing to note is that with only two exceptions, all the above studies use defaults to nudge participants. The exceptions are the study by Kroese et al. (2015), which used the reorganization of a check-out desk to steer impulse buying toward fruit instead of candy, and the study by Cheung et al. (2019), in which a cafeteria used two similar accessibility- and salience-based nudges, along with a social norm intervention.

Summarizing the state of knowledge

At present, there exist relatively much data on how people judge nudges in a description-based evaluation mode. Much less is known from judgments formed in an experience-based mode, and some studies of the latter kind have suffered from methodological limitations when tackling important issues. In relation to the framework presented earlier in this thesis this is unfortunate as an experience-based approach in most senses is better suited to inform the outlined principles. Centrally, the experience-based approach has the advantage that it allows for investigating actual experiences of nudge interventions, and actual experiences carry more weight than distanced opinions. For instance, in relation to informing ethical debates, Principle 1, a person’s experience (experience-based approach) of whether or not autonomy has been respected is more ethically relevant than a person’s projection about whether autonomy would be respected in a given situation (description-based approach). Similarly, an experience-based approach enables asking individuals

whether they feel happy, amused or helped by a nudge, which would constitute an increased net effect of the intervention, Principle 2, whereas a description-based approach is limited to asking participants whether they think they would feel better if they were to make a decision while subjected to the nudge. In other words, a description-based judgment is necessarily distanced, and there is reason to doubt how well it translates into real-world experience (see, e.g., Francis et al., 2016; Patil et al., 2014; Wilson and Gilbert, 2003). This is not to say that a description-based perspective is without merit. It is well suited for what it primarily is used for in this context: to investigate *opinion*. It can furthermore be argued that a description-based perspective has a benefit in that people may have an easier time to take a step back and reflect on their judgment when evaluating a nudge in this evaluation mode. My point here is simply that there is much more to people's thoughts of nudges than merely opinions. While I believe all principles outlined in the framework are essentially better informed by an experience-based rather than a description-based perspective, the optimal method in most cases is to use both approaches as different insights might be gained from each. It is surprising that mode of evaluation has not been paid more attention in research on people's views of nudges, particularly since the understanding that people's judgments and choices are malleable and susceptible to influence from situational factors forms a basic tenet behind many nudge interventions. Taking the above remarks into account, below follows a brief summary of where the state of knowledge stands in relation to each principle outlined in the framework.

Principle 1 suggests that people's perceptions and experiences of nudge interventions can inform ethical concerns with nudges. Previous literature conducted using a description-based approach provides many insights into which interventions people find acceptable and from which they fear intrusion to freedom of choice. However, there is an important lack of knowledge on the ethically central issue as to whether nudges find that nudges actually do prevent them from exercising autonomy. A few previous studies exist but were found to have important methodological limitations.

Principle 2 suggests that people's perceptions and experiences of nudge interventions themselves constitute a part of the nudge's net effect. Previous description-based studies provide clues as to how various nudges may be experienced in positive or negative ways. These perceptions are however necessarily distanced from any intervention's actual net effect, as the nudges only are judged in a hypothetical context. Some findings from experience-based research suggest common nudge interventions do not lower people's choice satisfaction, have little or no negative emotional influence, and do not evoke a perception of one's freedom of choice being threatened. However

these studies are limited from a lack of experimental control group, only using hypothetical choice tasks, and/or from measurement confounds.

Principle 3 suggests that people's perceptions and experiences of nudge interventions can moderate behavior change success, both at present and in the future. Assuming that a basic level of acceptability is a prerequisite for a nudge to successfully sway behavior, description-based research again has much to offer although in an indirect sense. Downstream behavioral consequences of how a nudge is experienced and perceived ought, however, to be researched using an experience-based approach. To my knowledge, only Wachner et al. (2020) have provided data on both choice experiences from an initial nudged choice, as well as data on a second behavioral outcome. In their study, autonomy, satisfaction, or pressure that was experienced in the nudged choice did not correlate with behavior in the second choice task. However, it remains to be investigated whether other perceptions or experiences, and other choice contexts, could fare differently.

Additionally, a few issues limit what conclusions can be drawn from previous research on the transparency of nudges. In the description-based research paradigm, several studies find that people prefer nudges that are transparent and target reflective cognitive processes, over nudges that are less transparent and target intuitive thinking or subconscious psychological processes (Davidai & Shafir, 2020; Felsen et al., 2013; Gold et al., 2020; Jung & Mellers, 2016). From these results, it is easy to leap to the inference that increasing the transparency of a nudge would make it be perceived more positively. However, the description-based studies to date have failed to investigate whether this is the case. These studies typically compare a transparent and a non-transparent intervention targeting the same behavior, but without keeping the intervention constant. For instance, in one study of smoking cessation interventions, designing cigarette packaging with off-putting graphic pictures and warning labels (i.e., a transparent nudge) was compared with making the filter a larger proportion of the cigarette (i.e., a non-transparent nudge). Here, more features than just the level of transparency of the intervention were changed. While justified in drawing inferences about groups of nudges, research in the description-based paradigm has so far failed to illuminate how increasing transparency on its own may influence perceptions of a nudge.

Studies using an experience-based approach have avoided the confound problem by manipulating transparency using explicit disclosures instead of integral features of the intervention. By introducing a nudge disclosure, however, new issues arise. The central issue, from which others are derivations, concerns what information is disclosed. At present, there is no general agreement among transparency researchers as to what information

should be disclosed. This has led to studies disclosing widely different levels of detail and contents about the intervention while purportedly addressing the same research question. From an ethical perspective, some examples may be considered insufficient to alleviate concerns of manipulation (e.g., disclosing that “we help you make healthier choices,” without mentioning the process how), while others may be considered over-sufficient and potentially manipulative themselves (e.g., “Please note the following: [by acting in line with the nudge], you guarantee that we will be able to accomplish our research objectives.”).

In the empirical studies of this thesis, to be described next, I aim to address several of the shortcomings and knowledge gaps identified above.

Summary of Empirical Studies

The three empirical studies in this thesis investigate how people perceive and experience default nudges and those who nudge them; how these perceptions are influenced by making the interventions more transparent; and whether this may affect subsequent choices in interactions with the same choice architect. The emphasis is on the experiences and perceptions of people who experience a default nudge first hand, instead of on the perceptions of people assessing descriptions of nudges. In the remainder, I will refer to the studies (the “papers”) using Roman numerals (I, II, and III), and to the sub-studies within these using Arabic numerals (1, 2, and 3).

The studies relate to the framework proposed earlier in this thesis. The framework suggested that people’s views of nudges are important for at least three reasons:

Principle 1: People’s perceptions and experiences of nudges inform the ethical assessment of nudges.

Principle 2: People’s perceptions and experiences of nudges are themselves part of the net effect of the nudge.

Principle 3: People’s perceptions and experiences of nudges can moderate present and future behavior change.

In addition, I suggested that a nudge’s level of transparency is relevant to its impact on the above concerns, and that the process of making a nudge transparent can, on its own, be relevant for the same reasons.

The empirical studies primarily target one or two specific parts of the framework. Study I focuses on how people experience being subjected to nudges, especially in relation to autonomy. It primarily addresses Principles 1 and 2. Study II focuses on how fair people perceive default nudges to be, and whether increased transparency influences fairness perceptions. This is examined in both description- and experience-based formats. It focuses primarily on Principles 1 and 2 as well as on transparency. Study III deals with how increasing the transparency of a default nudge influences perceptions of the choice architect, and whether this can influence subsequent choices in

interactions with the same choice architect. In relation to the framework, it focuses primarily on Principles 2 and 3.

In total, data from 5171 participants are used in the main experiments of these studies. All studies were conducted online with participants recruited from a panel hosted by Amazon, called Mechanical Turk (“MTurk”; www.mturk.com).

On open science practices

The last decade has been a tumultuous one for psychological science. Evidence from research on methodological practices (John, Loewenstein, & Prelec, 2012) and failed attempts to replicate the results of published research (e.g., Camerer et al., 2018; Klein et al., 2014, 2018; Open Science Collaboration, 2015) have led to concerns about the validity and robustness of large bodies of research. While the global structure of academic research is doubtless what upholds an environment conducive to sub-optimal research practices (Bakker, van Dijk, & Wicherts, 2012), the core issue nevertheless concerns decisions taken, intentionally or not, by individuals when conducting and reporting research. Flexibility in how data are collected, analyzed, and reported can severely increase the degree of false-positive results (Simmons, Nelson, & Simonsohn, 2011). A lack of transparency concerning the same aspects of the research process undercuts other researchers’ opportunities to properly assess, reproduce, and successfully replicate studies (O. Klein et al., 2018; Lindsay, 2020).

Adherence to methodological rigor is central to healthy science, especially when findings may be relevant to policy and used to inform interventions affecting people. In the empirical studies presented in this thesis, I have done my best to take steps in this direction. The data for all studies are openly available, as are the stimulus materials for all sub-studies. All sub-studies in studies II and III were preregistered⁴ (Nosek, Ebersole, Dehaven, & Mellor, 2018). That is, the study design, experimental conditions, measured variables, considerations for sample size, procedures of data collection, exclusions of participants, as well as all hypotheses and analysis plans (including analysis code specifying all hypothesis-testing analyses in exact detail) were determined and made public prior to data collection. In the event of any deviation from the preregistered plan, this was duly noted when reporting the study. Preregistration ensures that flexibility in analytic decisions, which normally risks being inadvertently used to support the researcher’s hypotheses,

⁴ I was not aware of the practice of preregistration and its merits when Study I was being conducted.

is kept to a minimum. It does not devalue exploratory research and analyses, but provides a convenient guide for differentiating what was predicted beforehand (and how this was to be tested) and what was found when exploring the data. Study III was additionally conducted as a registered report (Chambers, 2019), meaning that the theoretical rationale, study designs, hypotheses, and analysis plans were subjected to peer review before data collection, and thus benefitted from outside experts' suggestions in a proactive way.

Please note, however, that the reporting of the methods and results of the individual studies is somewhat shortened in this thesis. For full details, please see the original manuscripts found in the appendices.

Study I

The aim of Study I was to explore how people experience being subjected to nudges, with a focus on experiences and perceptions relevant to discussions of the ethics of nudging. The paper argues that: i) autonomy is a central aspect of the ethical concerns with nudging; ii) people's views as to whether autonomy is respected should inform this debate (but largely has not); and that iii) previous research has failed to assess autonomy in the most relevant sense by not subjecting people to nudges first hand.

We chose to study default nudges for being one of the most effective nudge interventions in influencing behavior (Hummel & Maedsche, 2019), and simultaneously perceived as among the most intrusive on freedom of choice (Hagman et al., 2015). Both features make default nudges of particular interest in relation to how well autonomy is respected. In three between-group experiments, we subjected participants (total $N = 2083$) to opt-out default nudges and measured their experiences of autonomy, choice-satisfaction, perceived threat to freedom of choice, and objection to the choice format (the last only in the third study). All studies compared choice experiences between participants subjected to an opt-out default nudge, an opt-in default, or an active choice format. Studies 2 and 3 additionally manipulated the level of transparency of the nudge by explicitly disclosing its presence and anticipated influence on behavior to participants. The results of all experiments were meta-analyzed and subjected to equivalence testing in Study 4.

Data and stimulus materials for all studies can be found at <https://osf.io/69be8/>.

Study 1: Experiencing a pro-environmental default nudge

Method

Participants ($n = 290$) faced a choice task in which they imagined moving into a new apartment (based on an experimental scenario in Steffel et al., 2016). The landlord would pre-fit the apartment with a number of appliances (e.g., dishwasher and refrigerator), and the participants' task was to decide whether they wanted the appliances in an environmentally friendly ("green") or a standard version. There were three experimental conditions. In the opt-out default condition, all appliances had been preselected as green, and participants had the choice to opt out of receiving each appliance in its green version (for a rent reduction of USD 1–7 per appliance). In the opt-in default condition, all appliances were instead preselected as standard, and participants had the choice to opt in to obtain them in their green versions (for a corresponding rent increase). In the active choice condition, nothing was preselected and participants had to indicate for each item which version they wanted to obtain. The opt-out default condition was considered the nudge condition, for encouraging the desired choices of green appliances. Subsequent to choosing appliance versions, participants answered questions about their experience of choosing and perceptions of the choice situation. Specifically, participants were asked about their experienced autonomy, choice satisfaction, and perceptions of threat to freedom of choice.

Results and discussion

Study 1 found that even though the opt-out format had a sizeable influence on choices, participants' experiences of autonomy, choice satisfaction, and perceived threat to freedom of choice did not significantly differ from those subjected to an opt-in or active choice format. In all conditions, participants reported high levels of experienced autonomy and choice satisfaction, and low perceptions of threat to freedom of choice. There were no significant differences between the experimental conditions. In other words, the findings were positive from the perspective of nudging: participants were successfully nudged by the default without reporting negative secondary outcomes.

This suggests that, when experienced first hand, the nudges were found to be less intrusive than reported in description-based studies, and less intrusive than assumed by critics. However, we reasoned that there were at least two other possible explanations of the results. One is that the participants might have failed to recognize the nudge and that it might influence their choices. A second is that since the choice task was hypothetical, and thus the lack of

experienced intrusion by the nudge could have been a result of insufficient engagement with the task. Studies 2 and 3 sought to shed light on these issues.

Study 2: Do choice experiences deteriorate when intervention transparency is increased?

Method

The second study used the same apartment acquisition scenario as in Study 1. Participants ($n = 606$) were again subjected to the choice of green versus standard appliances in one of three choice formats. However, Study 2 also manipulated the transparency of the intervention used. This was done by disclosing the nudge's presence and expected effect to half of the participants. Study 2 thus employed a 3(choice format: opt-out vs. opt-in vs. active choice) \times 2(transparency: disclosure present vs. absent) between-group experimental design. Specifically, the manipulation of level of transparency consisted of a text box, displayed to participants before they made their choices, disclosing: 1) that how a choice is formatted can influence people's choices, and 2) the direction in which the influence could be expected (i.e., low/average/high number of green appliances chosen). For instance, in the opt-out (nudge) condition, participants received information that preselecting an option makes the option more likely to be chosen, and that, in this context, this would lead to more green appliances being chosen.

Results and discussion

The results essentially mirrored those of Study 1. Participants in the opt-out default condition were again strongly influenced by the nudge in their choices of green appliances, and still reported high levels of experienced autonomy and choice satisfaction and low levels of perceived threat to freedom of choice. Surprisingly, experienced autonomy and choice satisfaction were even found to be significantly higher in the nudge condition than the opt-in default condition. However, the differences were small and should, in light of the other results in this paper, be interpreted with caution (see Study 4 below, which summarizes the findings).

There were no significant differences depending on level of transparency, that is, whether or not participants were given a disclosure of the choice architecture for any of the dependent variables. Nor were there any significant interaction effects between level of transparency and choice format.

Study 3: Do choice experiences deteriorate when choice stakes are increased?

Method

A limitation of studies 1 and 2 was that the experimental choice context was hypothetical and did not have real consequences for the participants. Without consequences arising from their choices, it could be that participants did not engage sufficiently with the task, which may have led to underestimation of the nudge's influence on participants' choice experiences. Study 3 ($n = 1187$) addressed this concern by introducing a monetary pay-off. Specifically, we replaced the hypothetical apartment scenario with a choice task in which participants decided between donating a bonus payment to charity or keeping it for themselves. After completing a filler task consisting of rating the similarity between geometrical shapes, participants were awarded a surprise USD 0.20 bonus payment, which could either be kept for oneself or donated to charity (i.e., a hurricane relief fund). The choice task was experimentally manipulated as in Study 2, in a 3(choice format: opt-out vs. opt-in vs. active choice) \times 2(transparency: disclosure present vs. absent) between-group design. The opt-out format here meant that donating the bonus was preselected on the screen, and would happen unless the participant actively made a change. Correspondingly, in the opt-in format, keeping the bonus for oneself was preselected. After making their choice of whether or not to donate the bonus payment, participants completed the same experience and perception measures as in the previous two studies. One question was added about whether the participant objected to how the choice was formatted.

Results and discussion

The pattern of results was similar to those in the previous two studies. Participants subjected to the opt-out default nudge were more likely to donate the bonus to charity, but their experienced autonomy and choice satisfaction were high, and perceived threat to freedom of choice and objection to the choice format low. The nudge's effect on donation choices was large, with 44% donating in the opt-out condition, compared with 26% in the opt-in and 36% in the active choice conditions. The differences found in Study 2 between opt-out and opt-in participants with regard to autonomy and satisfaction were not repeated, as no differences were found. Participants in the opt-out condition did, however, report a higher perceived threat to freedom of choice, although not a higher objection to the choice format, than did participants in the opt-in condition. This makes it seem as if the nudged participants were indeed able to

detect the attempt to influence, but, on average, did not mind this happening. Significant main effects of level of transparency were found for two dependent variables. For experienced autonomy, participants receiving a disclosure of the nudge's presence and expected effect gave somewhat lower ratings than did participants not receiving a disclosure. For perceived threat to freedom of choice, disclosed participants reported slightly higher levels.

Study 4: Internal meta-analysis and equivalence tests

Method

We concluded the paper by conducting aggregated analyses of all experience and perception variables from studies 1–3 (apart from the objection measure, which appeared only in the last study). The findings for experienced autonomy, choice satisfaction, and perceived threat to freedom of choice were meta-analyzed, and then subjected to equivalence testing in a TOST framework (Lakens, 2017; Lakens, Scheel, & Isager, 2018). Equivalence testing allows for statistical testing of the absence of an effect, as defined by the effect being smaller than a predefined smallest effect size of interest. In a TOST framework, this entails conducting two one-sided tests assessing whether the observed effect size is larger than a low equivalence bound, and smaller than a high equivalence bound. If both tests are significant, statistical equivalence, as defined by the chosen equivalence bounds, is declared. Ideally, equivalence bounds are based on practical or theoretical importance. For the present study, however, we lacked a rationale for deciding on any specific boundary, and no clear estimate from previous research existed. Therefore, we based our equivalence bounds on the statistical heuristic of what we had 80% power to detect at $\alpha = 0.05$ (as per recommendations in Lakens et al., 2018), which corresponded to an effect size of Cohen's d of ± 0.102 .

More precisely, we reasoned that the central issue was not whether the nudge led to *equivalent* experiences and perceptions, but whether the nudge worsened them to a meaningful degree. We thus focused on whether the nudge produced ratings not surpassing the “negative” equivalence bound. In other words, we tested whether the nudge led to experiences of autonomy and choice satisfaction that were significantly higher than the low equivalence bound (that is, whether the hypothesis that the difference was larger than $d = -.102$ could be rejected). Correspondingly, for perceptions of threat to freedom of choice, we tested whether the nudge led to ratings significantly lower than the high equivalence bound. In all comparisons, we tested participants in the opt-out default condition against participants in the active choice condition ($n = 1350$),

reasoning that the active choice condition was the most neutral control group, due to not being presented with any default course of action (comparisons with the opt-in default condition are reported in the paper's Supplementary Material). It also happened that for all tests, the active choice condition amounted to the most challenging test of the nudge condition.

Results and discussion

For all three variables, the differences between participants in the opt-out default condition and the active choice condition were small, or very small, according to conventional benchmarks (Cohen, 1988). Experienced autonomy (meta-analytic estimate: $d = 0.035$) and choice satisfaction (meta-analytic estimate: $d = 0.131$) for participants in the opt-out default condition were significantly higher than the low equivalence bound, establishing non-inferiority in comparison with participants in the active choice condition. We could, however, not reject the hypothesis that participants in the opt-out default condition perceived a higher threat to freedom of choice than did participants in the active choice condition (meta-analytic estimate: $d = 0.128$). The latter finding was driven by the results of Study 3.

In sum, we concluded that participants were able to recognize that a nudge aiming to change their behavior was in place, but that their experiences of choosing were not meaningfully diminished by this fact.

Study II

Study II sought to investigate how fair people find default nudges, and whether making the nudge more transparent could be a way to increase perceived fairness. If increased transparency would make a default nudge more acceptable to people, this could be of practical use to policymakers, as defaults tend to be more effective in influencing behavior than are other nudges (Hummel & Maedche, 2019), but simultaneously are perceived less favorably (Hagman et al., 2015). As noted earlier in the literature review, previous studies have found that people prefer transparent over non-transparent nudges, but it has not been investigated whether making one-and-the-same nudge transparent causes it to be perceived more favorably.

In three studies, participants (total $N = 1915$) rated the fairness of an opt-in default, a standard (not disclosed) opt-out default nudge, and a transparent (disclosed) opt-out default nudge. All studies used a donation choice similar to that of Study 3 in Study I, but with a pediatric cancer charity as the recipient

of the potential donation. A USD 0.20 surprise bonus payment was offered to participants and could either be donated to the charity or kept for oneself. In the opt-in default format, keeping the bonus was preselected. In the standard opt-out default format, donating the bonus was preselected. In the transparent opt-out default format, donating the bonus was preselected, and concurrently with the choice a text box was presented that disclosed the presence and expected influence of the default nudge.

Across the three studies, we varied the “evaluation mode” in which participants were asked to make their fairness assessments. That is, we varied whether participants rated the various choice formats based on descriptions of the nudge or after experiencing it themselves first hand, and whether in a joint or separate evaluation (only for the description-based evaluation). The details of these variations are described below in conjunction with the separate studies.

All studies were preregistered prior to the collection and analysis of data. The preregistrations, including analysis code and stimulus materials, and data for all studies can be found at <https://osf.io/fa9xy>.

Study 1: Description-based joint evaluation

Method

In Study 1, participants ($n = 439$) were tasked with rating choice formats in a description-based evaluation mode. The study instructions explained that there are multiple ways of formatting a choice task, and that we wanted their opinions of how fair two examples were in order to inform the design of future studies. Thus, participants were only shown (“described”), and rated, examples of choice formats; they were not themselves asked to engage in the choice task and make a decision as to whether or not to donate. There were two sets (“pairs”) of choice formats. Each participant viewed and rated one pair, with the two choice formats displayed simultaneously side by side. Pair 1 consisted of the opt-in default format and the standard opt-out default format. Pair 2 consisted of the standard opt-out default format and the transparent opt-out default format. The differences between the two formats were highlighted in yellow to facilitate comparison.

We hypothesized that the opt-in format would be perceived as fairer than the standard opt-out format, and that the transparent opt-out default format would be perceived as fairer than the standard opt-out format. After separately rating the fairness of each choice format in the pair, participants answered questions about what intentions they perceived that each format conveyed on

behalf of the choice architect. Specifically, participants were asked to what extent they agreed that the choice of choice format conveyed an intention to: influence the likelihood of donating; suggest a course of action; make choosing easy; and sufficiently inform the chooser.

Results and discussion

Both hypotheses were supported. When rated side by side, participants perceived the opt-in default format as fairer than the standard opt-out default format, and the transparent opt-out default format as fairer than the standard, non-disclosed opt-out format.

In addition, the results indicated that the three choice formats communicated different intentions of the choice architect. All intention measures were significantly different within each rated pair. A choice architect using a standard opt-out default, versus an opt-in format, was perceived to a higher degree as intending to suggest a course of action and to influence the decision maker, and to a lesser degree as intending to inform the decision maker sufficiently, and make choosing easy. The pattern was somewhat different between the standard and transparent opt-out default formats. Here, the transparent format was perceived to a higher degree as intended to influence the decision maker. This is interesting in relation to the fairness rating of the transparent opt-out default being higher. Presumably, the high level of perceived intended transparency (“inform the chooser sufficiently”) of the transparent format alleviated negative influences on the fairness measure.

Taken together, the results showed that participants found it fairer not to be nudged, but if they were to be nudged, that increased transparency made the nudge seem fairer. While the latter finding is positive for choice architects wanting to deploy defaults ethically, it should be remembered that the joint comparison format of Study 1 highlighted the differences between each presentation format in a way typically not found in real-world situations. Study 2 proceeded by investigating whether the benefit of increased transparency remained when interventions were evaluated separately.

Study 2: Description-based separate evaluation

Method

Study 2 closely resembled the first study, but the study design was changed from a within-subject to a between-subject design. Thus, instead of rating a pair of choice formats, participants were shown, and rated, only one format.

We viewed Study 2 as a robustness check of the results of Study 1: would differences between choice formats remain when the opportunity to directly compare formats had been removed? For our analyses, we posited the same hypotheses as in Study 1: that the opt-in format would be perceived as fairer than the standard opt-out format and, similarly, that the transparent opt-out default format would be perceived as fairer than the standard opt-out format.

Results and discussion

As suspected, the results indicated that differences between choice formats were smaller in a between-subject design. Hypothesis 1 was still supported, as the opt-in default format was again perceived as fairer than the standard opt-out default format. However, we found no support for Hypothesis 2, as no difference in perceived fairness was found between the standard and transparent opt-out default formats. A similar pattern was found for the measures of perceived intentions behind the choice format selection. The opt-in default was significantly different from both opt-out conditions for all intention measures, although differences between the standard and transparent opt-out formats largely disappeared. It seems that without a reference point, any benefits of the transparent format were invisible to participants.

Study 3: Experience-based separate evaluation

Method

Study 3 presented participants ($n = 1052$) with the choice that had merely been described in studies 1 and 2. That is, the participants here engaged in the task first hand, and actually made the decision as to whether to donate the bonus payment to charity. In addition to fairness perceptions, we thus also looked at participants' likelihood of donating the money. Regarding choice, we hypothesized that participants in the standard and transparent opt-out default conditions, respectively, would donate more than participants in the opt-in default condition. We further hypothesized that donations in the two opt-out conditions would be statistically equivalent (with equivalence bounds set at a ± 0.08 difference in proportions) in a TOST analysis (Lakens, 2017; Lakens et al., 2018). Study 3 did not include the measures of perceived intentions of the choice architect.

Results and discussion

For perceptions of fairness, the pattern of results differed from that seen in the previous two studies. The opt-in default was no longer rated as significantly fairer than the standard opt-out default (even though it was still descriptively the fairest). Unexpectedly, the transparent opt-out default was also perceived as significantly *less* fair than either the opt-in or standard opt-out default format.

For choice, significant default effects were found in that participants in both the standard and transparent opt-out default format conditions to a higher extent chose to donate their bonus to charity (47% and 53%, respectively) than did participants in the opt-in default condition (26% donated). We did, however, not find support for statistical equivalence between the two opt-out default conditions. Notably, this was because participants in the transparent condition donated too much, and not too little.

The finding that transparency led to a decrease in perceived fairness is ironic, in that the purpose of a disclosing a nudge in the first place is to make it more ethical. It was also surprising given the results of the previous two studies, as well as other research (e.g., Paunov et al., 2019a). We interpret this finding as indicating that increasing the transparency of a nudge can be a double-edged sword, in that as well as communicating positive openness, a disclosure could highlight that a negatively interpreted influence attempt is being made. That we only saw a negative effect of the disclosure in the third study may be due to the first-hand evaluation mode leading to higher engagement with the choice. Possibly only when experienced, and with money on the line, did the perceived intended influence indicated in previous experiments become strong enough to actually decrease fairness ratings.

Study III

Study III had two main aims: i) to investigate how disclosing a default nudge, at various points in time, influences a nudgee's perception of the choice architect (nuder), and ii) to investigate to what extent perceptions of the choice architect affect choices in present or future interactions with the choice architect. In line with Principle 3 of the framework presented in the introduction, we reasoned that to the extent that the nudgee views providing a disclosure of the nudge positively, this could lead to the nudgee wanting to reciprocate the favor by acting compliantly at a later stage. Specifically, Study 1 investigated whether disclosure of a default nudge before or after making a

choice influences perceptions of the choice architect and one's willingness to stand by the choice when given the chance to change. Study 2 investigated whether the timing of the nudge disclosure (before or after choice) and perceptions of the choice architect influence willingness to comply with a subsequent request by the choice architect.

Study III was conducted as a registered report (Chambers, 2019), which means that the study was subjected to peer review, and eventually accepted for publication, before data were collected. Preregistrations and code for all confirmatory analyses, stimulus materials, and data for all studies can be found at <https://osf.io/463af/>.

Study 1: Perceptions of the choice architect and timing of nudge disclosure

Method

In the choice task of Study 1, participants ($n = 415$) were asked whether they, after the present study was finished, minded completing an unpaid additional questionnaire to provide data for a student research project. All participants were subjected to an opt-out default nudge in that agreeing to complete the additional questionnaire was made the preselected alternative. All participants were also presented a disclosure of the default nudge, making its presence and anticipated effect on choice transparent. However, the timing with which the disclosure was presented was varied between the two experimental groups. In the pre-disclosed condition, participants were shown the disclosure concurrently with choosing, meaning before they had made up their mind. In the post-disclosed condition, participants were given the same disclosure after having made the choice. Specifically, in the survey flow, the choice task was followed by a measure of participants' perceptions of the choice architect (i.e., us, the researchers). Participants in the post-disclosed condition were presented with the disclosure after completing this task. At the same point in the survey, pre-disclosed participants were shown a reminder of the disclosure. Next, participants in both conditions were given the opportunity to revise their responses, regarding both perceptions of the choice architect and whether they wanted to complete the additional unpaid questionnaire. Thus, all participants gave two ratings of the choice architect, and stated twice whether or not whether they wanted to complete the additional questionnaire. The scale measuring perceptions of the choice architect consisted of seven items, and asked to what extent participants agreed that we, the choice architects,

appeared fair, honest, trustworthy, open, manipulative, deceptive, and controlling.

Our main predictions were that post-disclosed, but not pre-disclosed, participants' perceptions of the choice architect would deteriorate from measurement 1 to measurement 2, and that post-disclosed participants who initially agreed to complete the additional questionnaire would be more likely to change their choice when asked the second time. We also tested, in an exploratory way, whether being presented with the disclosure at the time of making the choice the first time affected decisions.

Results and discussion

Participants' perceptions of the choice architect deteriorated between the first and second measurements. To our surprise, this happened in both the pre- and post-disclosed conditions. However, in line with our hypothesis, the tendency to give lower ratings was stronger for post-disclosed participants. The decrease did not amount to more than about one quarter of a scale point (0.13 scale point in the pre-disclosed condition), and ratings were still on the positive side of the scale for both conditions. Furthermore, there was no difference between conditions at the first measurement point, when one condition had a nudge disclosure and the other did not.

The hypothesis that post-disclosed participants who had agreed to the request in the first choice would be more likely to change their decision when asked a second time was not supported. All in all, only 10 out of 415 participants changed their initial choice when prompted the second time. We also found no significant difference in initial choices between pre- and post-disclosed participants (or at the second prompt either). Participants in both conditions acted highly prosocially, with 66% agreeing to complete the additional unpaid questionnaire in the pre-disclosed condition and 73% in the post-disclosed condition.

The results of Study 1 indicated that choice architects who fail to disclose their use of nudges risk being perceived less favorably, if people find out about the nudge later on. While choice architects may not normally plan to disclose a nudge after the fact, it is far from unlikely that a nudgee may find out about the intervention from other sources, such as peers or the media. Reassuringly from the perspective of choice architects, the decrease did not lead to participants revoking their initial agreement to cooperate with the choice architect's request. However, this is not to say that punishing the choice architect in subsequent behavior could not ensue in other situations. For instance, this could happen if acting in line with the nudge demanded a higher effort or cost and the decision maker felt tricked into engaging in the behavior.

Study 2: Perceptions of the choice architect and subsequent behavior

Method

Study 2 ($n = 758$) used the same basic design and materials but with a few additions and modifications. The main choice task was the same as in Study 1, and subsequent to making their choice, participants responded to the same measure of perceptions of the choice architect. However, perceptions of the choice architect were measured at only one point. In addition, the reaffirmation of whether participants wanted to complete the additional questionnaire was replaced with a new, separate, second choice task. Specifically, after responding to the first choice task and stating their perceptions of the choice architect, all participants were asked whether they minded providing their email address in order to be contacted about unpaid participation in a future research study.

There were four experimentally manipulated between-group conditions. As in Study 1, participants in the pre-disclosed condition received a nudge disclosure concurrently with making the first choice, and participants in the post-disclosed condition received the same disclosure after having chosen (but before stating their perceptions of the choice architect). In addition, we included a third condition in which the nudge was never disclosed to participants. Lastly, there was a fourth condition in which the default alternative in the first choice task was to reject completing the additional questionnaire (i.e., an opt-in default format). This enabled us to examine whether or not the default to agree to complete the questionnaire in the other three conditions had successfully nudged participants' choices. The reject-default condition did not include a disclosure of the choice format.

We posited a number of hypotheses regarding participants' choices and perceptions of the choice architect. For the first choice, the main prediction was that there would be a default effect, such that participants with the default set to agree to complete the additional questionnaire would be more likely to do so than participants with the default set to reject. We were also interested in exploring whether disclosure of the nudge affected choices. For perceptions of the choice architect, we predicted that participants would give lower ratings when given the disclosure after choosing, as compared with before choosing. For the second choice, we predicted that participants in the post-disclosed condition would provide their email address less frequently compared with the pre-disclosed condition. We were also interested in exploring whether the data

would be consistent with a mediation model, such that the disclosure would influence perceptions of the choice architect that in turn would influence participants' behavior in the second choice task.

Results and discussion

As expected, whether the default was set to agree to or reject completing the additional questionnaire (Choice 1) influenced participants' choices. In the three conditions in which the default was to agree with the request, 71–75% of participants did so. In comparison, when the default was to reject completing the additional questionnaire, 58% agreed. Whether or not the nudge was disclosed to participants in Choice 1 did not significantly influence their choice.

For perceptions of the choice architect we found, as in Study 1, that participants to whom the nudge had been disclosed after the fact provided lower ratings. Again, ratings were still on the positive side of the scale in all conditions.

Participants in the post-disclosed condition did not, however, as had been predicted, agree with the second request (Choice 2) to a lesser degree. Nor did any of the other conditions stand out in Choice 2, as agreement levels in all conditions ranged between 29–34%. Thus, even though participants' perceptions of the choice architect were significantly related to agreement with the second request, we did not find support for the proposed mediation model.

The findings were highly consistent with what was found in Study 1, and the results of Study 1 were extended in two major ways. First, Study 2 evidenced that participants were successfully nudged by the default in Choice 1. Second, we tested whether the influence of the nudge or disclosure had downstream behavioral consequences, as materialized in Choice 2, and found no evidence that it did.

General Discussion

The aim of this thesis was to investigate how people experience and perceive default nudges. This was motivated by three reasons outlined in a framework presented in the thesis' introduction. The framework was developed with an aim to address criticisms suggesting that nudges are insufficiently effective in influencing behavior and are unethical by infringing on people's autonomy. In the following, I first synthesize the findings of the empirical studies and relate them to previous research. Next, methodological strengths and weaknesses of the present studies are discussed. Thereafter, I consider theoretical and practical implications of the present studies, and relate the findings to the framework for appreciating the value of people's perceptions and experiences of nudges that was developed in the introduction. Finally, directions for future research are suggested.

Taking stock of people's perceptions and experiences of default nudges

Choice experiences

In the introduction, I argued that especially people's *experiences* when subjected to a nudge intervention are relevant to assessing the ethics of nudging (Principle 1) as well as to assessing the net effect of an intervention (Principle 2). Study I investigated two choice experiences: participants' experienced autonomy and their choice satisfaction.

In relation to the ethical assessment of nudging, I concluded that respect for autonomy is at the heart of concerns about the use of nudges. I argued that if we take seriously the *as judged by themselves* criterion, meant to privilege the nudgee's assessment of the outcomes of a nudge intervention, then central ethical questions concerning nudging are empirical (also, in more modestly empirical epistemological accounts, the data are informative). I then subsequently argued that most research on nudges and autonomy is limited from using an indirect form of measurement, a "description-based approach," in which participants are detached from any choosing and therefore are limited

to speculation concerning control in decision-making. The few studies that measured autonomy after subjecting participants to nudges first hand were found to have important caveats.

Study I addressed the main shortcomings identified in the previous studies (e.g., small samples, lack of transparency, and lack of a baseline comparison group) and provided data from three experiments. The results indicated that experienced autonomy was consistently high among, and not reduced, for participants subjected to the default nudge. The robustness of this finding was strengthened by the fact that the pattern of results did not change when participants were given a disclosure of the presence and expected effect of the nudge, or when the choice concerned a monetary decision. In other words, it could be ruled out that participants were simply unaware of the nudge and its effect, and that the results extended to a choice with real consequences for the participants.

The results of Study I align well with the findings of Abhyankar et al. (2014) and Wachner et al. (2020), who also found reports of high experienced autonomy from participants subjected to default nudges. With respect to Abhyankar et al.'s (2014) study, the present results extend its findings in several ways. First, we removed a possible confound in the autonomy measure. In their study, the focus was on a broader choice process that contained more components than only the nudge, such as a second non-nudged reassessment of the choice. It is therefore impossible to establish what influence on autonomy may have stemmed from the nudge, and what may have stemmed from other parts of the setup. Second, we ensured that participants were aware of the nudge by presenting a nudge disclosure. Third, we included a choice with real consequences and, fourth, we recruited a vastly larger total sample. With respect to Wachner et al. (2020), Study I centrally extends their findings by showing that experienced autonomy was *not reduced* for participants subjected to the default nudge. This is important for ethical reasons, as even though experienced autonomy was high among nudged participants, the nudge could still have made them feel less autonomous.

Study I's results are, however, in conflict with the result emphasized by Arvanitis et al. (2020), namely, that default nudges are detrimental to experienced autonomy. As discussed in the introduction (p. 20), I believe there are grounds for some caution concerning this result. In essence, Arvanitis et al.'s (2020) study was conducted with a relatively small sample, and the decrease in autonomy appeared to have low convergent validity as judged by their other autonomy measures. Therefore, their results appear limited in scope or may lack robustness. Thus, in light of the compatibility with other previous research, and given the benefits of a larger sample size (approximately 15 times larger) together with methodological improvements (e.g., disclosure of the

nudge and use of a consequential choice), I suggest it is more reasonable to favor the findings of Study I. That is, in light of the currently available empirical evidence, people subjected to common instances of default nudges do not experience their autonomy to be lowered as a consequence of the nudge.

In addition to experienced autonomy, Study I examined how participants' satisfaction with their choice was affected by being subjected to a default nudge. As for autonomy, the findings indicated consistently high ratings and no deteriorating effect of the nudge. These results are consistent with those of Abhyankar et al. (2014) and Wachner et al. (2020), and relate to these studies in the same way as discussed above for experienced autonomy.

Perceptions of the nudge and of the choice architect

All three studies of this thesis provided insights into how people perceived the nudge and the choice architect. The overarching story told by the empirical findings is that people who experience default nudges first hand do not object to the nudges (Study I), find the use of the nudge to be fair (although somewhat less fair than an opt-in default) (Study II), and perceive the nudge as posing little threat to freedom of choice (Study I). A choice architect using a default nudge is perceived favorably (Study III), even though participants infer an intention to influence their choice (Study II).

The findings align well with previous results from experience-based studies of people's perception of nudges, which similarly have found perceptions to be positive or unconcerned (see pp. 22–23). In the literature review, I suggested that when evaluated in an experience-based evaluation format, people seemingly report more favorable perceptions of the nudge than in a description-based format. This pattern was also indicated in the present studies. Primarily, Study II found that when experienced first hand (Study 3), the nudge received higher fairness ratings than in a description-based format (studies 1 and 2). However, caution is warranted regarding this conclusion, as the comparisons are between studies whose scale end-points differed somewhat, something which may have affected the absolute levels of the ratings. Nonetheless, findings of Study I suggest that the difference between description- and experience-based paradigms may be real, in that perceptions of threat to freedom of choice, and objections to choice format, were markedly lower than for similar interventions in description-based research (e.g., Hagman et al., 2015). In sum, this suggests that when actually experiencing a nudge first hand, people deem the nudge less of a threat than they would have expected.

Choices

All three studies of this thesis consistently found strong default effects. That is, designating a course of action as the default by preselecting a choice option led to this option being chosen more often. These findings add to a growing body of evidence that defaults have a pervasive effect on decision making, and that default effects are robust in a wide range of applications (Jachimowicz et al., 2019). Primarily, the present studies contribute to knowledge of default effects by showing that they prevail when disclosed to decision makers, as elaborated on below.

Transparency

The thesis's main findings in relation to transparency concern choices, evaluation modes, and timing of the disclosure.

In relation to effects on choices, six out of six experiments found that disclosing the nudge to nudgees did not significantly influence their choices. This pattern is consistent with the growing body of studies finding no, or negligible, influence on choice (see pp. 21–22). The present studies contribute by obtaining this result in high-powered experiments, containing consequential choice tasks, and with demonstrably attentive participants who either displayed understanding of the disclosure (Study I) or passed an elaborate general attention check⁵ (studies II and III). Additionally, no evidence was found that a nudge disclosure presented in a first choice influenced participants' behavior in a second choice (Study III).

Moreover, Study II showed that increased transparency, by means of a nudge disclosure, can affect how fair the nudge is perceived to be. Interestingly, we found that the effect could go in both directions, either positively or negatively, depending on the circumstances of the evaluation. When evaluated side by side with a non-transparent version of the same nudge in a description-based format, a disclosed nudge was perceived as fairer. However, when experienced first hand (without a comparison point) in an actual choice, the same disclosed nudge was perceived as less fair than its undisclosed counterpart. This result nuances the findings of previous description-based studies that people prefer transparent to non-transparent types of nudges, by showing that making one and the same nudge transparent does not straightforwardly improve how it is perceived.

⁵ An unpublished pilot study showed that approximately 85% of participants passing the Instructional Manipulation Check (Oppenheimer et al., 2009) we used also passed a comprehension check for the disclosure contents.

In relation to the timing of a nudge disclosure, Study III found that the choice architect was perceived less favorably when the nudge was disclosed after the choice. Similarly, Steffel et al. (2016, Experiment 2b) had previously documented that participants perceived the choice architect as more ethical when they disclosed the nudge before rather than after the choice. With the addition of an undisclosed baseline condition in the second study of Study III, we are able to show that perceptions of the nudge or choice architect were instead worsened by the post-disclosure rather than improved by the pre-disclosure.

Methodological limitations and considerations

Sample considerations

All of the present studies recruited participants from the online panel Amazon Mechanical Turk (MTurk; www.mturk.com). The use of online panels such as MTurk has increased dramatically in popularity during the last decade (Anderson, Allen, Plante, Quigley-McBride, Lovett, & Rokkum, 2019). However, the appropriateness of these panels has sometimes been called into question (e.g., Anderson et al., 2019; Hauser, Paolacci, & Chandler, 2018). Skeptics have feared that online participants are inattentive and that the data quality is low. Yet, empirical support for such claims appears mixed, with some noting that MTurk data quality often compares favorably to data obtained using other common sampling methods (Hauser et al., 2018). The issue of trying to judge online participant pools is complicated due to high turnover rates among panelists (Jain, Darma, Parameswaran, & Wisdom, 2017). This means that what may hold true at one point can quickly change over time. For instance, the onset of the COVID-19 pandemic, during which some of the data for this thesis were collected, appears to have changed the pool of MTurk participants (Arechar & Rand, 2021).

In any case, the use of MTurk for this thesis has provided several benefits. A first one is that I have been able to recruit proper sample sizes to study effects the sizes of which are common in experimental cognitive and social psychology. As described earlier, in the meta-analytic equivalence tests of the fourth study of Study I, we could reliably detect effect sizes corresponding to a Cohen's d of 0.10. Assuming normality of distributions, this corresponds to being able to reliably detect a statistically significant difference between two distributions sharing a 96% overlap. For reference, the sometimes-used rule of thumb to collect 20 participants per experimental condition (cf. the since regretted recommendation by Simmons et al., 2011, p.1363) only enables one

to, with the same power of 0.80, detect differences between two groups corresponding to a Cohen's d of 0.91. Only one between-group comparison of a nudge perception or experience variable in this thesis would have reached statistical significance with this number of participants (i.e., perceived fairness for the opt-in vs. standard opt-out default, in Study 2 of Study II).

A second benefit is that I have been able to access a much wider demographic than traditional recruitment bases such as student samples would have allowed. This considerably broadens the generalizability of the findings. Nevertheless, highly "WEIRD"⁶ samples (Henrich, Heine, & Norenzayan, 2010) were used in the current studies. It cannot be ruled out that a broader set of participants would have yielded different results. For example, Henrich et al. (2010) reviewed evidence that judgments of fairness can differ between cultures. Consequently, the results of Study II might have turned out differently if the research had used a non-WEIRD sample. Cultural differences in perceptions of nudges have previously been documented in description-based studies (Sunstein et al., 2017). The lack of research on non-WEIRD samples in this thesis is a limitation impeding the practical use of the findings, as behaviorally informed policy interventions are by no means limited to WEIRD countries but are being used around the globe (see, e.g., World Bank, 2018).

A third benefit is that conducting the studies in an online environment, without ever meeting or directly communicating to the participants other than through the questionnaire, likely diminished the risk of participants' responses being affected by social desirability. Had data collection taken place face to face, participants may have been more likely to answer in ways thought wanted by the experimenter.

Participant's level of engagement with the studies and choices

A related concern is whether participants were sufficiently engaged with the experimental tasks we asked them to undertake. Were our nudge manipulations sufficiently strong that negative experiences and perceptions were likely to show themselves? There are at least two reasons to believe that they were.

First, based on earlier findings in the survey literature on perceptions of nudges, the types of nudges used in this thesis are the ones most likely to provoke people. As mentioned several times throughout this thesis, default nudges are perceived as the type of nudge that is most intrusive to freedom of

⁶ WEIRD is an acronym for Western, Educated, Industrialized, Rich, and Democratic. Research has shown that samples drawn from societies with such characteristics can sometimes differ in meaningful ways from samples from other parts of the world. Therefore, empirical findings derived from WEIRD samples do not necessarily generalize well to other parts of the world (Henrich et al., 2010).

choice, especially when imposing losses on decision-makers (Hagman et al., 2015; Sunstein & Reisch, 2019). All three studies in this thesis used defaults to nudge participants, and in applications with real consequences for the participants in the form of money or effort lost. Since people routinely express objection to disliked circumstances out of principle, even when stakes may appear small, it is reasonable to expect that participants would have objected in at least some of these settings—should really have found the nudge intrusive when experienced first hand.

Second, while the choice stakes in the consequential choices were modest, it is still likely that they signaled a meaningful level of value to participants. As detailed in footnote 3 to Study I (see Appendix 1, pp. 19-20), we have found in a pilot study that participants were willing to work on a menial task, at a sub-par pay rate, for several minutes in order to keep a bonus payment of the amount offered in studies I and II.

Additionally, we used demanding attention and comprehension checks to screen the samples. In Study I we assessed comprehension of the nudge disclosure, which a surprising amount of participants failed, and verified the main results in sub-group analyses containing only participants passing the check (reported in the original article, see Appendix I). In studies II and III, we used an instructional manipulation check (a “trap question”; Oppenheimer et al., 2009) to exclude inattentive responders. The instructional manipulation check was designed by us for the specific study materials and could thus not be recognized from other studies. Especially in Study III, a large proportion of the sample failed this check (see Study 2 in the original article, Appendix III). That large groups of participants failed the attention and comprehension checks suggests the checks were relatively demanding screening instruments, and thus that the remaining participants were attentive and engaged with the experimental tasks.

Limits and merits of introspection

Another objection that could be raised is that it is difficult to introspect about a choice experience, such as autonomy or satisfaction. This concern primarily affects Study I. A skeptic could, for instance, object that nudges may have distorted views of their own autonomy, and argue that the value of such a measurement therefore is limited. My response is that I think we ought to acknowledge people’s responses at face value when their level of awareness of the circumstances is as high as it was in the present studies. In particular, in Study I, the results were essentially unchanged in the sub-group analyses containing only participants disclosed of the nudge and passing a comprehension check for this information. If these participants are considered

subjectively autonomous or satisfied only in an inauthentic way, I think it is sensible to ask whether we are using the same yardstick for assessing these experiences that we are using for people in other everyday situations. It seems that, ordinarily, we do accept someone's claim that they are happy and acknowledge that this mental state is valuable, regardless of whether an observer believes that the judgment lives up to some objective standard.

This is not to say that there are not limits to what even "successful" introspection can tell us. As noted in the introduction concerning autonomy and ethics (pp. 11–12), it is perfectly possible to deny that a subjective sense of autonomy settles or even informs ethical assessments. Appeals to subjective experience may be countered with the "happy slave"-objection, arguing that one's circumstances can be undesirable even though subjectively perceived as acceptable. I recognize that the present findings may conceivably be construed as an argument *against* acknowledging subjective experience when assessing the ethicality of nudging. If it is taken for granted on theoretical grounds that the default nudges in the present studies did infringe on autonomy, then the findings merely go to show that people consistently fail to realize their misfortune.

Nevertheless, the aim of the present studies has been descriptive and not normative. I would argue that Study I shows that people were sufficiently aware of their circumstances that their level of introspection was a meaningful measurement. As far as empirical data are considered informative, I therefore believe Study I can help illuminate discussions of ethics (i.e., autonomy measurement) and the net effect of nudge interventions (i.e., choice satisfaction measurement).

Methodological approach

The research questions addressed in this thesis have primarily concerned comparisons of choices, perceptions, and experiences between participants subjected or not subjected to nudges. Given this aim, I believe that the use of experimental research designs has been appropriate, and even necessary. Nevertheless, how people perceive and experience nudges could be studied in more ways. In particular, I believe a qualitative approach, for example, by interviewing people, would have been a valuable undertaking. Perceptions and experiences, especially of complex matters such as autonomy, manipulation, and fairness risk being conveyed in simplified and un-nuanced ways when measurement scales restrict what responses can be given. People evidently have lots of opinions of nudges and nudging, and a broader and deeper understanding of these thoughts could likely be reached by allowing people to speak more freely. To my knowledge, at present only one study by Junghans,

Cheung, and De Ridder (2015) has explored the same research theme using a qualitative approach.

Staying within an experimental approach, studies in naturalistic environments, especially where the researcher is not also transparently playing the role of the choice architect, would additionally be desirable to conduct. I would not agree, however, that the present studies are at fault for having been conducted in online questionnaires, since the majority of them still concern real, consequential, choices. I perceive engaging with these research themes fully out of the lab to be a natural continuation of the present studies.

Generalizability

The ways in which a nudge can be designed, and the contexts in which it can be applied, are limited only by the inventiveness of the world's choice architects. Furthermore, as noted in the introduction, due to definitional uncertainty it can be hard to judge what is and is not considered a nudge. I acknowledge that the present findings certainly cannot account fully for this heterogeneity and cannot appropriately answer for perceptions and experiences of nudges in a fully general sense. Rather, we might always need to take a piecemeal approach to studying nudging (Wilkinson, 2013).

That said, the choice contexts used in this thesis are relatively similar to common everyday cases in which defaults are used (e.g., paying a small fee for green instead of standard shipping). Also, survey research shows that default nudges are deemed the most intrusive type of nudges, especially when they incur losses (Hagman et al., 2015; Sunstein & Reisch, 2019). Since we did not find very negative effects on people's experiences and perceptions when applying default nudges in such situations, there is some reason to expect "milder" types of nudges to fare better than the ones tested here.

Ethical considerations

To participate in the present studies, participants were required to complete a consent form. The consent form disclosed the broader topic of the study, the monetary compensation offered, the expected duration of the study, and that data would be treated confidentially and stored for a minimum of ten years. The level of personal detail in the data collected, including regarding demographic matters, was kept low. Participants were free to refuse to answer all demographic questions. Payment levels for study participation were set according to established norms for short studies on the recruitment platform we used, and approximated minimum wage in the country where participants resided (the USA).

Still, the research questions concerned the alleged manipulation of people, and it was necessary to subject participants to nudges in order to examine these questions. Whether or not this entailed a level of manipulation that participants were uncomfortable with was ironically one of the substantive issues researched. As discussed, our findings indicate that participants were at least subjectively unaffected in this regard. We furthermore took steps to prevent or minimize manipulation that could harm participants. For instance, participants were never nudged to give away their agreed-on payment for participating in the study. In all studies involving monetary choices, the money that a nudge encouraged participants to donate was money that they had received as an unexpected bonus during the study. When choices concerned spending additional time and effort, that is, completing an extra measurement scale or signing up on an email list, these tasks were kept short and as convenient as possible. Approximately half of all recruited participants were additionally informed that the nudge they were subjected to in the choice task could influence their decisions, and in what direction this was anticipated to occur.

At another level of ethical consideration, all studies were conducted and reported with the aim of maximizing the transparency of the research process, facilitating scrutiny of the results, and supporting future replication attempts. For all studies, data and stimulus materials are openly available under a Creative Commons 4.0 CC-BY license. Studies II and III openly share code scripts to enable reproduction of the results of all main analyses. In addition, all sub-studies in studies II and III had their participant recruitment plans, materials, hypotheses, and analysis plans preregistered prior to data collection. This was done to protect against liberal use of “researcher degrees of freedom” (Simmons et al., 2011) that might unintentionally have led to distortions of the results.

On theory and practical implications

The present studies were largely descriptive and exploratory in nature. This is a result mainly of two factors. First, the specific research questions examined have often been motivated by normative issues, such as ethical concerns, rather than by psychological theory. As a result, hypotheses have often not been posed to test theoretically derived empirical predictions, but instead to provide data to illuminate theoretical discussions and debates. Second, at the start of this thesis project, research on how people perceive and experience nudges consisted of little more than a handful of studies. Consequently, I perceived that there were many basic research questions open to examination, and that there existed no specific theory of how people perceive and experience nudges

from which to derive and test ideas (however, Hagman, 2018, subsequently formulated a theory on the public acceptability of nudges).

As discussed in the introduction, the concept of nudging is notoriously ubiquitous, and integration efforts have resulted mainly in list-based frameworks rather than unifying theory (at least in any concrete sense of the word). I believe that attempts at theorizing how people perceive and experience nudges risk running along a similar path, with scattered micro-theories for separate types of interventions. However, the situation will become clearer over time as more empirical research is conducted, and it is possible that general patterns will emerge. For instance, a nudge's level of coercion and its level of facilitating choosing may be central factors that affect many nudges' perceptions and experiences.

Until that point, I hope that the framework developed in this thesis can help in systematizing how perceptions and experiences of nudges can be conceived, and function as an argument for the value of acknowledging them. In relation to the three principles outlined in this framework, the following are the main points that have been learned from the empirical work in this thesis.

Principle 1: People's perceptions and experiences of nudges inform the ethical assessment of nudges

I identified two senses in which perceptions and experiences can be ethically informative: first, to let the voices of those affected by nudges be heard and, second, to assess the empirical feasibility of the premises or conclusions of normative arguments made by ethicists. The empirical work of this thesis provides findings regarding both senses.

Concerning the first aspect, I showed that people evaluating a default nudge found the nudge to be fair, when evaluated in both description- and experience-based evaluation formats (Study II). While the nudge was perceived as fair, participants consistently perceived a non-nudge, i.e., an opt-in default, as a fairer arrangement of the same choice task. In addition, participants reported low levels of objection and perceived threat to freedom of choice for the default nudges they encountered (Study I).

Concerning the second aspect, I found that contrary to common criticisms of nudges (and especially default nudges), participants subjected to default interventions did not experience low or lowered autonomy (Study I).

Taken together, these findings indicate that people appear to accept default nudges, especially when they experience the nudge first hand. Importantly, fears of nudges infringing on autonomy appear exaggerated, at least as judged by people themselves.

A practical implication of these findings is that policymakers may rest assured that the people affected by a default nudge will not consider the intervention unethical.

Principle 2: People's perceptions and experiences of nudges are themselves part of the net effect of the nudge

Against the backdrop of concerns that nudges are manipulative interventions, it is positive that there was no, or negligible, evidence of negative perceptions or experiences on the part of people subjected to default nudges in the present studies. Had being subjected to the nudge led to frustration or resentment, this would have lowered the net effect of the intervention. Perceptions of listed under the above heading, i.e., positive fairness, low objection, and low perceived threat to freedom of choice, are all positive in this sense, as are the findings of the nudge not reducing choice satisfaction (Study I) and of positive perceptions of the choice architect (Study III).

In fact, in the meta-analyses of Study I's results, participants subjected to the default nudge were actually significantly more satisfied with their choice than other participants were. Technically, this adds an additional positive benefit to the net effects of those interventions. However, the meta-analytic effect size was small ($d = 0.13$) and was insufficient to surpass what was considered a negligible effect in that study. I therefore currently hesitate to overemphasize this result.

Taken together, the present findings did not provide evidence that taking a broader approach to assessing the net effect of the nudges, by including people's perceptions and experiences of being nudged, led to a less favorable result. Instead, some tentative evidence was found for favorable results.

A practical implication of these findings is that policymakers may not be able to expect additional side-effects from default nudge interventions in the form of improved perceptions or experiences. However, this will necessarily vary by application context, and I believe it is likely that defaults can provide a benefit by making choosing easier in situations in which people risk facing information overload (which was not investigated in the present studies). Other nudges may additionally provide benefits such as making choices more fun for people.

Principle 3: People's perceptions and experiences of nudges can moderate present and future behavior change

Study III found that participants who perceived the choice architect more favorably were more likely to behave in line with the nudge and agree to the choice architect's requests. Similarly, in Study I, participants' choices correlated negatively with perceived threat to freedom of choice and with objection to the choice format. That is, participants who perceived a low threat to freedom of choice, or did not object to the choice format, were more likely to engage in the behavior encouraged by the nudge. However, in neither of these cases is the direction of causality established. While it appears plausible that participants having positive perceptions would therefore be more likely to act in line with the nudge (as Principle 3 suggests), it is also possible that acting in line with the nudge led to more favorable ratings, for instance, as a justification of their behavior. Yet another possibility is that both the positive perceptions and behavior was driven by a third factor, such as a general prosocial sentiment. Furthermore, Study III also found that participants' perceptions of the choice architect correlated positively with their likelihood of agreeing with a subsequent request from the same choice architect. A caveat is warranted in that in none of the present studies was it found that being subjected to the nudge moderated the relationship between participants' perception and behavior. That is, it was not shown that the nudge played an "active" part in the link between perceptions and behavior; rather, the same relationships were found for participants who were not subjected to the nudge.

A practical implication of these findings is that policymakers who desire effective nudges would likely do well to design interventions that are perceived favorably by the people subjected to them. More research is needed, however, to clarify the causal relationship at work.

A general note on practical implications

Concerning the practical implications of the present findings more generally, it is appropriate to once again stress that nudges are a highly heterogeneous class of interventions. Outcomes may differ depending on the specifics of the intervention design, and outcomes may vary between different applications of one and the same nudge. It is therefore my primary recommendation that policymakers and others who apply nudges should routinely assess how people perceive and experience the nudge before it is implemented at a large scale, and use this information to guide the design of their interventions. The framework developed in this thesis can act as a guide for how perceptions and experiences can be considered.

Suggestions for future research

Good science progresses incrementally, from researchers repeatedly assessing the robustness of phenomena in new and different circumstances. For research that aims to inform policy, potentially at a large scale, this is especially important. When under-researched behavioral science phenomena are used as a basis for policy interventions, this risks wasting public resources and, in the worst case, even having harmful impacts on those affected. Related to tackling the COVID-19 pandemic, there has been debate on the state of behavioral science evidence and its potential to inform large-scale policy (IJzerman et al., 2020; van Bavel et al., 2020). Critics have argued that insights from behavioral science are often unfit to inform large-scale interventions.

Among other issues, it has been argued that behavioral science phenomena have commonly been evidenced by only a narrow range of stimuli, been assessed using under-validated measures (and with only WEIRD samples), been estimated with low precision, and too often fail to replicate (IJzerman et al., 2020). Regardless of whether these claims are true of behavioral science in general, they certainly at present apply to the sub-field of how people perceive and experience nudges. Due to the rarity of such research, at least using what I call an experience-based approach (in which people engage with nudges first hand, not only read descriptions of them), I would suggest that more basic research is currently needed. The present findings, such as those on experienced autonomy, need to be replicated and assessed in new circumstances if they are to properly inform policy.

Future studies should furthermore explore whether other types of nudges, such as norm interventions, produces similar results. Most of the experimental designs used in this thesis would be readily adaptable to use with other types of nudges, and in other application contexts. Another issue is to what extent, and under what circumstances, people perceive that potentially negative choice experiences may be justified by their behavioral outcomes.⁷ Interesting cases may arise when people have conflicting preferences, such as simultaneously desiring and not desiring to smoke. Other topics for future research may be how perceptions and experiences of being nudged may be moderated by ease of opting-out (i.e., resistibility), by people's preferences for the behavior they are being nudged toward, and by people's attitudes toward the choice architect (cf. Tannenbaum et al., 2017).

As mentioned, I think that field studies, in which the agent employing the nudge is someone other than the researcher, would be important to pursue in future research. There is a risk that the current findings may be biased because it was the researcher who both employed the nudge and posed the questions

⁷ I thank Carl Martin Allwood for this suggestion.

about the intervention; for instance, social desirability may play a role in the results.

As also noted, I further believe that a broader methodological approach is necessary. A deeper and more nuanced understanding of how people perceive and experience nudges could be gained by using qualitative methods such as interviews.

Concluding remarks

In this thesis, I have argued that it is important to take seriously people's perceptions and experiences of nudges, and have developed a framework outlining what I believe are the major ways in which this is true. In particular, I have suggested that research on people's perceptions and experiences of nudges can inform debates on the ethics and effectiveness of nudging, especially when taking into account the perspective of people subjected to nudges first hand. I believe that a first-hand perspective is generally preferable to what I have called a description-based perspective, as in most aspects it is ultimately what is actually experienced that is central to informing ethical and effectiveness concerns.

The results of the present empirical studies suggest that common default nudges meet important ethical standards, as judged by the people subjected to them, in that the nudges were not considered a threat to autonomous choice and were perceived to be fair. In relation to effectiveness, understood as a broadening of what should count toward the net effect of a nudge, neither especially positive nor negative results were found. In a sense, this might still be viewed as a positive result for nudging, if evaluated against the fear that nudges are manipulative and could cause resentment toward the choice architect or spark contrarian behavior.

The empirical studies in this thesis are among the first, but will certainly not be the last, inquiries into the perceptions and experiences of people subjected to nudges. As I have emphasized in this concluding discussion, I believe more research is necessary before any general conclusions should be attempted. Nevertheless, I think that the present studies ever so slightly tilt the evidence base further in the direction of a favorable verdict on default nudges. Based on the currently available evidence, I suggest that the default position should be to consider default nudges a truly libertarian paternalistic form of intervention, in that defaults can successfully encourage behavior change while leaving autonomous decision-making intact.

References

- Abhyankar, P., Summers, B., Velikova, G., & Bekker, H. (2014). Framing Options as Choice or Opportunity: Does the Frame Influence Decisions? *Medical Decision Making*, *34*(5), 567-582.
- Al-Ubaydli, O., Lee, M., List, J., Mackevicius, C., & Suskind, D. (2021). How can experiments play a greater role in public policy? Twelve proposals from an economic model of scaling. *Behavioural Public Policy*, *5*(1), 2-49. doi:10.1017/bpp.2020.17
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9-10), 1082-1095.
- Allcott, H., & Kessler, J. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, *11*(1), 236-276.
- Anderson, C., Allen, J., Plante, C., Quigley-Mcbride, A., Lovett, A., & Rokkum, J. (2019). The MTurkification of Social and Personality Psychology. *Personality and Social Psychology Bulletin*, *45*(6), 842-850.
- Arechar, A. A., & Rand, D. G. (2021). Turking in the time of COVID. *Behavior Research Methods*, 1-5.
- Arimura, T., Li, S., Newell, R., & Palmer, K. (2012). Cost-effectiveness of electricity energy efficiency programs. *Energy Journal*, *33*(2), 63-99.
- Arvanitis, A., Kaminiotis, K., & Kalliris. (2019). Are defaults supportive of autonomy? An examination of nudges under the lens of Self-Determination Theory. *Social Science Journal*, 1-11.
- Bakker, M., Van Dijk, A., & Wicherts, J. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*(6), 543-554.
- Bang, H. M., Shu, S. B., & Weber, E. U. (2020). The role of perceived effectiveness on the acceptability of choice architecture. *Behavioural Public Policy*, *4*(1), 50-70.

- Behavioural Insights Team (2014). *EAST: Four simple ways to apply behavioural insights*.
- Benartzi, S., Beshears, J., Milkman, K., Sunstein, C., Thaler, R., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8), 1041-1055.
- Bergquist, M., Nilsson, A., & Schultz, W. P. (2019). A meta-analysis of field-experiments using social norms to promote pro-environmental behaviors. *Global Environmental Change*, 59, 1-18.
- Bovens, L. (2009). The ethics of nudge. In Grüne-Yanoff T. & Hansson S.O. (Ed.), *Preference change* (pp. 207-220). Dordrecht: Springer.
- Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Luistro Jonsson, M., & Rahali, B. (2018). Can nudges be transparent and yet effective? *Journal of Economic Psychology*, 65, 41-59.
- Bruns, H., and Perino, G. (2019), 'The role of autonomy and reactance for nudging-experimentally comparing defaults to recommendations and mandates' *Pre-print* retrieved from: <https://ssrn.com/abstract=3442465>
- Cadario, R., & Chandon, P. (2018). Which healthy eating nudges work best? A meta-analysis of field experiments. *Appetite*, 130, 300-301.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637-644.
- Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism". *University of Pennsylvania Law Review*, 151(3), 1211-1254.
- Carlsson, F., Gravert, C., Johansson-Stenman, O., & Kurz, V. (2019). *Nudging as an Environmental Policy Instrument*.
- Chambers, C. (2019). The registered reports revolution Lessons in cultural reform. *Significance*, 16(4), 23-27.
- Chapman, G., Li, M., Colby, H., & Yoon, H. (2010). Opting in vs opting out of influenza vaccination. *JAMA*, 304(1), 43-44.
- Cheung, Tracy T. L., Gillebaart, Marleen, Kroese, Floor M., Marchiori, David, Fennis, Bob M., & De Ridder, Denise T. D. (2019). Cueing

- healthier alternatives for take-away: A field experiment on the effects of (disclosing) three nudges on food choices. *BMC Public Health*, 19(1), 1471-2458.
- Clavien, C. (2018). Ethics of nudges: A general framework with a focus on shared preference justifications. *Journal of Moral Education*, 47(3), 366-382.
- Cohen, J. (1988), *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.
- Conly, S. (2013), *Against autonomy: Justifying coercive paternalism*, Cambridge University Press.
- Coons, C., & Weber, M. (2013). *Paternalism: Theory and practice*. Cambridge: Cambridge University Press.
- Coons, C., & Weber, M. (2014). *Manipulation: Theory and practice*. New York: Oxford University Press.
- Davidai S., & Shafir, E. (2020). Are ‘nudges’ getting a fair shot? Joint versus separate evaluation. *Behavioural Public Policy*, 4(3), 273-291.
- Davison, W. P. (1983), The Third-Person Effect in Communication. *The Public Opinion Quarterly*, 47(1): 1-15.
- Delaney, L. (2020, June 16). Ethics of behavioural science policy: Nudge FORGOOD. *Economics, psychology, policy*. <http://economicpsychologypolicy.blogspot.com/2015/11/ethics-of-nudging-bunch-of-reading.html>
- DellaVigna, S., & Linos, E. (2020). *RCTs to Scale: Comprehensive Evidence from Two Nudge Units*.
- Dinner, I., Johnson, E., Goldstein, D., & Liu, K. (2011). Partitioning default effects: Why people choose not to choose. *Journal of Experimental Psychology*: 17(4), 332-341.
- Dranseika, V., & Piasecki, J. (2020). Transparent defaults and consent for participation in a learning health care system: An empirical study. *Journal of Empirical Research on Human Research Ethics*, 15(4), 261-270.
- Dworkin, G. (2013). Defining paternalism. In C. Coons and M. Weber (Eds.) *Paternalism* (pp. 25-38). Cambridge University Press.
- Egebark, J., & Ekström, M. (2016). Can indifference make the world greener? *Journal of Environmental Economics and Management*, 76, 1-13.

- Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: public attitudes towards overt and covert nudges. *Judgment and Decision Making*, 8(3), 202-213.
- Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action?, *PloS one*, 11(10),
- Gold, N., Lin, Y., Ashcroft, R., & Osman, M. (2020). ‘Better off, as judged by themselves’: Do people support nudges as a method to change their own behavior? *Behavioural Public Policy*, 1-30.
- Goodwin, T. (2012). Why We Should Reject ‘Nudge’. *Politics*, 32(2), 85-92.
- Hagman, W. (2018). *When are nudges acceptable? Influences of beneficiaries, techniques, alternatives and choice architects*. Doctoral dissertation, Linköping University Electronic Press.
- Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public views on policies involving nudges. *Review of Philosophy and Psychology*, 6(3), 439-453.
- Hansen, P. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, 7(1), 155-174.
- Hansen, P., & Jespersen, A. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, 4(1), 3-28.
- Halpern, D. (2015). *Inside the nudge unit: How small changes can make a big difference*.
- Hausman, D. M., and Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy*, 18(1), 123-136.
- Hauser D, Paolacci G, Chandler JJ. (2020) Common concerns with MTurk as a participant pool: Evidence and solutions. In Kardes FR, Herr PM, Schwarz N, (eds.). *Handbook of research methods in consumer psychology*. New York: Routledge
- Hedesström, M., Michaelsen, P., Nyström, L., Luke, T. J., & Johansson, L-O. (2019). What’s the net benefit of a nudge? Exploring behavioral spillover from choosing a default. *Manuscript in preparation*.
- Hedlin, S., and Sunstein, C. R. (2016). Does active choosing promote green energy use: Experimental evidence. *Ecology LQ*, 43, 107-141.

- Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83.
- Horvath J., & Wiegmann, A. (2021). Intuitive expertise in moral judgments. *Australasian Journal of Philosophy*, DOI:10.1080/00048402.2021.1890162
- House of Lords, Science and Technology Select Committee. (2011). Behaviour Change: Second report of Session 2010-12. London, UK
- Hummel, D. & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80, 47-58.
- Ivanković, V., & Engelen, B. (2019). Nudging, transparency, and watchfulness. *Social Theory and Practice*, 45(1), 43-72.
- Ito, K. (2015). Asymmetric incentives in subsidies: Evidence from a large-scale electricity rebate program. *American Economic Journal: Economic Policy*, 7(3), 209-237.
- Jain, A., Sarma, A. D., Parameswaran, A., & Wisdom, J. (2017). Understanding workers, developing effective tasks, and enhancing marketplace dynamics. *Proceedings of the VLDB Endowment*, 10(7), 829-840.
- 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*, 1-28.
- Jaeger, C., & Schultz, P. (2017). Coupling social norms and commitments: Testing the underdetected nature of social influence. *Journal of Environmental Psychology*, 51, 199-208.
- John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524-532.
- John, P. (2018). *How far to nudge? : Assessing behavioural public policy*. Edward Elgar Publishing: Cheltenham, UK.
- Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649), 1338-1339.
- Jung, J. Y., & Mellers, B. A. (2016). American attitudes toward nudges. *Judgment & Decision Making*, 11(1), 62-74.

- Junghans, A., Cheung, T., & De Ridder, D. T. D. (2015). Under consumers' scrutiny - an investigation into consumers' attitudes and concerns about nudging in the realm of health behavior. *BMC Public Health*, *15*(1), 336.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291.
- Kauppinen A. (2014) Ethics and Empirical Psychology – Critical Remarks to Empirically Informed Ethics. In Christen M., van Schaik C., Fischer J., Huppenbauer M., Tanner C. (eds) *Empirically Informed Ethics: Morality between Facts and Norms*. Springer, Cham. https://doi.org/10.1007/978-3-319-01369-5_16
- Klein, R. A., Ratliff, K., Vianello, M., Adams, R., Bahník, &, Bernstein, M., . . . Nosek, B. (2014). Investigating variation in replicability: A “Many Labs” replication project. *Social Psychology*, *45*(3), 142-152.
- Klein, R. A., Vianello, M., Hasselman, Fred, Hasselman, B., Tybur, J.M., Van Lange, P.A.M., . . . A-Lab. (2018). Many labs 2: Investigating variation in replicability across sample and setting. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443-490.
- Klein, O., Hardwicke, T., Aust, F., Breuer, J., Danielsson, H., Mohr, A., & Ijzerman, H. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1).
- Kosters, M., & Van Der Heijden, J. (2015). From mechanism to virtue: Evaluating nudge theory. *Evaluation*, *21*(3), 276-291.
- Krijnen, J., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy*, *3*(2), 1-18.
- Kroese, F. M, Marchiori, D. R, & De Ridder, D. T. D. (2016). Nudging healthy food choices : A field experiment at the train station. *Journal of Public Health*, *38*, E133-3842.
- Ijzerman, H., Lewis, N., Przybylski, A., Weinstein, N., Debruine, L., Ritchie, S., . . . Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, *4*(11), 1092-1094.
- Lacko, P. & Nilsson, G. (1988). Bilbältesanvändning i Sverige 1983-1986 [Car seat belt use in Sweden 1983-1986]. *VTIrapport*, *326*, 1-23.
- Lades, L. & Delaney, L. (2019). Nudge FORGOOD. *Behavioural Public Policy*: 1–20.

- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355-362.
- Lakens, D., Scheel, A.M., & Isager, P. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269.
- Larrick, R., & Soll, J. (2008). The MPG illusion. *Science*, 320(5883), 1593-1594.
- Lindsay, D. (2020). Seven steps toward transparency and replicability in psychological science. *Canadian Psychology*, 61(4), 310-317.
- Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2015). Warning: You are about to be nudged. *Behavioral Science & Policy*, 1(1), 35-42.
- Loewenstein, G., & O'Donoghue, T. (2006). "We can do this the easy way or the hard way": Negative emotions, self-regulation, and the law. *University of Chicago Law Review*, 73(1), 183-206.
- Lovakov, A., & Agadullina, E.R. (2021), Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*,
- Löfgren, Å., & Nordblom, K. (2020). A theoretical framework of decision making explaining the mechanisms of nudging. *Journal of Economic Behavior & Organization*, 174, 1-12.
- Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics*, 116(4), 1149-1187.
- McKenzie, C. R., Liersch, M. J., and Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, 17(5): 414-420.
- Münscher, R., Vetter, M., & Scheuerle, T. (2016). A review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making*, 29(5), 511-524.
- Nilsson, A., Erlandsson, A., Västfjäll, D., & Tinghög, G. (2020). Who are the opponents of nudging? Insights from moral foundations theory. *Comprehensive Results in Social Psychology*, 1-34.
- Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and social psychology bulletin*, 34(7), 913-923.

- Nosek, B., Ebersole, C., Dehaven, A., & Mellor, D. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2600-2606.
- Nys, T. R., and Engelen, B. (2017). Judging nudging: Answering the manipulation objection. *Political Studies*, 65(1): 199-214.
- OECD (2019), *Tools and Ethics for Applied Behavioural Insights: The BASIC Toolkit*, OECD Publishing, Paris.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349.
- Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., and Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, 9(1): 94-107.
- Paunov, Y., Wänke, M., & Vogel, T. (2019a). Transparency effects on policy compliance: disclosing how defaults work can enhance their effectiveness. *Behavioural Public Policy*, 1-22.
- Paunov, Y., Wänke, M., & Vogel, T. (2019b). Ethical defaults: Which transparency components can increase the effectiveness of default nudges? *Social Influence*, 14(3-4), 104-116.
- Paunov, Y., Wänke, M., & Vogel, T. (2020). Combining defaults and transparency information to increase policy compliance. *Social Psychology*, 51(5), 354-359.
- Pe'er, E., Feldman, Y., Gamliel, E., Sahar, L., Tikotsky, A., Hod, N., & Schupak, H. (2019). Do minorities like nudges? The role of group norms in attitudes towards behavioral policy. *Judgement and Decision Making*, 12(1), 40-50.
- Pichert, D., & Katsikopoulos, K. V. (2008). Green defaults: Information presentation and pro-environmental behaviour. *Journal of Environmental Psychology*, 28(1), 63-73.
- Rebonato, R. (2012), *Taking liberties: A Critical Assessment of Libertarian Paternalism*, Palgrave Macmillan.
- Reisch, L., & Sunstein, C. (2016). Do Europeans like nudges? *Judgment and Decision Making*, 11(4), 310-325.
- Schmidt, A., & Engelen, B. (2020). The ethics of nudging: An overview. *Philosophy Compass*, 15(4).

- Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.
- Smith, C., Goldstein, D., & Johnson, E. (2013). Choice without awareness: Ethical and policy implications of defaults. *Journal of Public Policy & Marketing*, 32, 159-172.
- Steffel, M., Williams, E., & Pogacar, R. (2016). Ethically deployed defaults: Transparency and consumer protection through disclosure and preference articulation. *Journal of Marketing Research*, 53(5), 865-880.
- Sunstein, C. R. (2014). *Why nudge? The politics of libertarian paternalism*. New Haven, CT: Yale University Press.
- Sunstein, C. R. (2015). *Choosing not to choose: Understanding the value of choice*, Oxford University Press
- Sunstein, C. R. (2016). *The ethics of influence: Government in the age of behavioral science*. Cambridge University Press.
- Sunstein, C. R. (2018). Misconceptions about nudges. *Journal of Behavioral Economics for Policy*, 2 (1), 61-67.
- Sunstein, C. R., Reisch, L., & Rauber, J. (2017). A worldwide consensus on nudging? Not quite, but almost. *Regulation & Governance*, 12(1), 3-22.
- Sunstein, C. R., & Thaler R. H. (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review*, 70, 1159-1667.
- Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, 31(3), 355-366.
- Tannenbaum, D., Fox, C. R., & Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour*, 1(7).
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New York, NY: Penguin Books.
- Thunström, L. (2019). Welfare effects of nudges: The emotional tax of calorie menu labeling. *Judgment and Decision Making*, 14(1), 11-25.

- Yan, H., & Yates, J. F. (2019). Improving acceptability of nudges: Learning from attitudes towards opt-in and opt-out policies. *Judgment and Decision Making*, 14(1), 26-39.
- Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., . . . Kitayama, S. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5), 460-471.
- Wachner, J., Adriaanse M., and De Ridder, D. T. D. (2020). The influence of nudge transparency on the experience of autonomy, *Comprehensive Results in Social Psychology*, 1-16.
- White, M. (2013). *The manipulation of choice: Ethics and libertarian paternalism*, Springer.
- Wilkinson, T. (2013). Nudging and manipulation. *Political Studies*, 61(2), 341-355.
- Wilson, T., & Gilbert, D. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, 14(3), 131-134.