

Nudging in actie

Beleidsexperimenten ontwerpen op basis van gedragsinzichten

Pieter Raymaekers, Ellen Fobé, Koen Migchelbrink, Amandine Lerusse & Marleen Brans

D/2021/10106/014

Vlaanderen
is vernieuwend bestuur

STEUNPUNT
BESTUURLIJKE
VERNIEUWING

Het Steunpunt Bestuurlijke Vernieuwing is een consortium van:

KU LEUVEN
INSTITUUT VOOR DE OVERHEID

Parkstraat 45 bus 3609
B-3000 Leuven
Tel: 0032 16 32 32 70

Prof. dr. Geert Bouckaert
Promotor-coördinator
geert.bouckaert@kuleuven.be

Prof. dr. Annie Hondeghem
Verantwoordelijke KU Leuven
annie.hondeghem@kuleuven.be

Politics & Public Governance
Universiteit Antwerpen

Sint-Jacobstraat 2
B-2000 Antwerpen
Tel: 0032 3 265 53 87

Prof. dr. Wouter Van Dooren
Verantwoordelijke UAntwerpen
wouter.vandooren@uantwerpen.be

200 JAAR
UNIVERSITEIT
GENT

Campus Mercator G
Henleykaai 84
B-9000 Gent
Tel: 0032 9 243 29 04

Prof. dr. Joris Voets
Verantwoordelijke UGent
joris.voets@ugent.be

UHASSELT

Martelarenlaan 42
B-3500 Hasselt
Tel: 0032 11 26 81 11

Prof. dr. Johan Ackaert
Verantwoordelijke UHasselt
johan.ackaert@uhasselt.be

Steunpunt Bestuurlijke Vernieuwing
Parkstraat 36 bus 3606 - 3000 Leuven - België
Tel: 0032 16 32 36 10 - E-mail: sbv@kuleuven.be - URL: https://steunpuntbestuurlijkevernieuwing.be/

© Steunpunt Bestuurlijke Vernieuwing
Niets uit deze uitgave mag, zelfs gedeeltelijk, openbaar gemaakt worden, gereproduceerd, vertaald of aangepast, onder enige vorm ook, hierin begrepen fotokopie, microfilm, bandopname behoudens uitdrukkelijke en voorafgaande toestemming van de uitgever.
Dit rapport vermeldt de mening van de auteur en niet deze van de Vlaamse overheid.
De Vlaamse overheid kan niet aansprakelijk worden gesteld voor het gebruik dat wordt gemaakt van de in deze bekendmaking opgenomen gegevens.

STEUNPUNT
BESTUURLIJKE
VERNIEUWING

/ rapport

NUDGING IN ACTIE

BELEIDSEXPERIMENTEN ONTWERPEN OP BASIS VAN GEDRAGSINZICHTEN

27.10.2021

Pieter Raymaekers, Ellen Fobé,

Koen Migchelbrink, Amandine Lerusse & Marleen Brans

INHOUD

Inleiding 8
1 Het gedragskundig perspectief 11
1.1 Een vijfde beleidsperspectief? 11
1.2 Nudging in perspectief 17
1.3 Het functionele perspectief 18
2 De experimentele methode 22
2.1 Wat is een experiment? 23
2.2 Welk type experimenten? 26
2.3 Waarom experminenteren in bestuur en beleid? 29
2.4 Hoe experimenteren? Een stappenplan 34
2.5 Waar liggen de experimentele grenzen? 38
3 Survey-Experiment Snelheidsboetes 41
3.1 De gedragsfactor in het verkeer 44
3.1.1 Protection Motivation Theory 44
3.1.2 Nudging en rijsnelheid 46
3.2 Rationele versus emotionele boodschappen 48
3.2.1 De rationele interventie 49
3.2.2 De emotionele interventie 50
3.3 Design en operationalisatie 52
3.3.1 Drie brieven, drie groepen 52
3.3.2 Intentievragen en PMT constructen 54
3.4 Data-analyse en resultaten 55
3.4.1 Balansverhoudingen 56
3.4.2 Effect op de intentie om zich aan de snelheidslimiet te houden 57
3.4.3 Effect op de drijfveren van verandering in intentie via de Protection Motivation Theory 60
3.4.4 Appreciatie van de rationele en emotionele boodschap 62
3.4.5 Lessen leren en repliceren: het pre(lude)-experiment 65
3.5 Conclusies 66
3.6 Aanbevelingen 68
4 Veldexperiment parkeerboetes 70
4.1 Scope: Van GAS tot GASAM 73
4.1.1 GAS boetes voor fout parkeren en stilstaan 73
4.1.2 GASAM procedures 75
4.1.3 GASAM cijfers 77
4.2 Nudging en belastingen 79
4.2.1 Afschrikken, herschikken of op de moraliteit mikken? 80
4.2.2 Britse veldexperimenten 82
4.2.3 Belgische veldexperimenten 85
4.3 Nudging en parkeerboetes: interventies 87
4.3.1 Interventie 1: Maak de boodschap gemakkelijk en vereenvoudig de brief 89
4.3.2 Interventie 2: Maak de boodschap opvallend en accentueer mogelijke extra kosten 90
4.3.3 Interventie 3: Maak de boodschap sociaal en beschrijf wat de meeste mensen doen 91

4.4 Een geclusterd design 92
4.4.1 Het bepalen van de betalingsgraad 92
4.4.2 Van batches naar clusters 93
4.5 Resultaten 95
4.5.1 Geobserveerde proporties 95
4.5.2 Balansverhoudingen 96
4.5.3 Significantietesten 97
4.5.4 Experimenteren = lessen leren 100
4.6 Conclusies 102
4.7 Aanbevelingen 103
5 Besluit 105
6 Referenties 109

FIGUREN

Figuur 1. Schematische voorstelling van een klassiek experiment met een pretest posttest control group design 23
Figuur 2. Het basisprincipe van een Randomized Controlled Trial 24
Figuur 3. Aantal gepubliceerde experimenten in de top 20 tijdschriften over bestuurs- en beleidskunde 1990-2015 30
Figuur 4. Type experimenten 31
Figuur 5. Type onderzoekseenheden 31
Figuur 6. Protection Motivation en Speeding Behaviour 46
Figuur 7. Notice of Intended Prosecution en emotionele bijlage 48
Figuur 8. Eerste deel van de rationele bijlage 49
Figuur 9. Tweede deel van de rationele bijlage 50
Figuur 10. Emotionele bijlage in deze studie 51
Figuur 11. Boetebrief met rationele en emotionele bijlage voor het randomized survey-experiment 53
Figuur 12. Aandeel deelnemers met appreciatiescore 7/10 of meer in de twee interventiegroepen 63
Figuur 13. Percentage deelnemers dat 'helemaal' of 'eerder' akkoord gaat met de appreciatievragen 64
Figuur 14. Verkeersbord parkeerverbod (E1) 75
Figuur 15. Logo GASAM 76
Figuur 16. Procedures GASAM 77
Figuur 17. Aantal GASAM dossiers parkeren en stilstaan per jaar (2015-2020) 78
Figuur 18. Percentages tijdige betalingen per jaar (2015-2019) 79
Figuur 19. Checklist voor brieven en mails Vlaamse Overheid 88
Figuur 20. Maak de boodschap gemakkelijk en vereenvoudig de brief 89
Figuur 21. Maak de boodschap opvallend en accentueer mogelijke extra kosten 90
Figuur 22. Maak de boodschap sociaal en beschrijf wat de meeste mensen doen 92
Figuur 23. Percentage op tijd betaalde brieven per interventie (op basis van geobserveerde proporties) 96
Figuur 24. Resultaten interventies in odds ratios (binnen 95% betrouwbaarheidsinterval) 99

TABELLEN

Tabel 1. Vijf perspectieven op het beleidsproces 16
Tabel 2. Experimentele designs 25
Tabel 3. Kenmerken per type experiment 29
Tabel 4. Opmaak van groepen in het randomized survey-experiment 53
Tabel 5. Balance checks: verhoudingen tussen de groepen op basis van enkele parameters 56
Tabel 6. Verhoudingen tussen mannen en vrouwen in de drie groepen volgens hun leeftijd 57
Tabel 7. Intentie algemeen, vergeleken in de voormeting en de nameting 58
Tabel 8. Intentie in voormeting en nameting weergegeven per groep en per snelheidsregime 59
Tabel 9. Effecten van de boodschappen op de PMT constructen 61
Tabel 10. Verdeling groepen naar geslacht en leeftijd in pre-experiment 65
Tabel 11. Interventiezinnen en resultaten eerste veldexperiment VK 83
Tabel 12. Interventiezinnen en resultaten tweede veldexperiment VK 84
Tabel 13. Interventiezinnen en resultaten veldexperiment België 86
Tabel 14. Aantal respondenten en clustergroottes 95
Tabel 15. Balanstesten 97
Tabel 16. Resultaten interventies in odds ratios 100

BOXEN

Box 1. Problemen bij het implementeren van beleidsexperimenten 39
Box 2. Samenvatting case rijsnelheid volgens stappenplan 43
Box 3. Samenvatting case parkeerboetes volgens stappenplan 72

INLEIDING

Bij de meeste maatschappelijke uitdagingen waar overheden zich voor geplaatst zien, spelen gedrag en gedragsbeïnvloeding een cruciale rol. Wil een overheid dat burgers hun belastingen correct en op tijd aangeven, de verkeersregels respecteren, gezonde en duurzame levenskeuzes maken en zich massaal laten vaccineren tegen COVID-19, dan zal ze telkens rekening moeten houden met het bewuste én het onbewuste keuzegedrag van burgers. Naast het sturen van gedrag via traditionele beleidsinstrumenten zoals regels, subsidies, accijnzen of communicatiecampagnes duiken plots nieuwe werktuigen op in de toolbox van beleidsmakers. Zo maken overheden steeds meer gebruik van gedragswetenschappelijke inzichten in het beleidsproces, met nudging als het meest uitgesproken voorbeeld (Leong & Howlett, 2020). Op nationaal en internationaal niveau is een belangrijke rol weggelegd voor zogenaamde *behavioural insights teams* en *nudge units* (Ball & Head, 2020; Feitsma, 2019b)

Binnen beleidsonderzoek en -praktijk ontwaren we een tweevoudige beweging: een eerste richting een nieuw soort inzichten in beleid en een tweede richting een nieuwe aanpak voor het testen van beleid. Er is, ten eerste, sprake van een '*behavioural turn*' waarbij theorieën en modellen uit disciplines zoals sociale psychologie en gedragseconomie circuleren op de tekentafels van beleidmakers (Strassheim & Beck, 2019; Sunstein, 2020). Deze omslag zet de rationale keuzetheorie die de burger als rationele, weloverwogen beslisser opvoert onder druk. Het beeld van de *homo economicus* wordt uitgedaagd door en aangevuld met het beeld van de *homo psychologicus*, dat rekening houdt met emoties, routines en het onbewuste beslissingsproces van mensen (Battaglio et al., 2019; Leong & Howlett, 2020; Tummers, 2019b). Niet enkel de gedragswetenschappelijke modellen en theorieën, maar ook de methoden uit de sociale psychologie en gedragseconomie worden overgenomen. Tegelijk met de '*behavioural turn*' zijn we, ten tweede, immers getuige van een '*experimental turn*' waarbij overheden en onderzoekers experimenten opzetten om causale verbanden te identificeren en te testen welke beleidsacties effectief zijn (James, John, et al., 2017; John, 2017c; Van de Walle, 2017).

Beide tendensen duiken simultaan op, staan in onderling verband, maar hoeven niet noodzakelijk samen te vallen. Onderzoek naar beleidsthema's zonder uitgesproken gedragsdimensie kan zich eveneens bedienen van een experimenteel design (Hansen & Tummers, 2020). Evengoed kunnen gedragsinterventies onderzocht worden via andere methoden en technieken zoals casestudies, diepte-interviews of focusgroepen (John,

2018). Toch zien we dat gedragsinterventies en nudges steeds vaker worden ontworpen en geëvalueerd vanuit een experimentele logica, op basis van pilots en zogenaamde *Randomized Controlled Trials* (Einfeld, 2019; John, 2020).

De bestuurskundige analyse achter de opmars van gedragsinzichten en experimenten in beleid werd reeds uitvoerig belicht in een vorig rapport: "Nudging in perspectief. Een verkennend kader voor de toepassing van gedragsinzichten in beleid" (Raymaekers et al., 2019). In dit rapport zetten we de theoretische bevindingen uit dat literatuuronderzoek om in concrete acties . We bieden een antwoord op de volgende vraag: *Hoe ontwerp je een beleidsexperiment op basis van gedragsinzichten?*

Ondertussen bereikten ons vanuit Nederland de resultaten van het multidisciplinaire onderzoeksproject "Welfare Improvement through Nudging Knowledge" (WINK) waarin psychologen, ethici en bestuurskundigen de krachten bundelden om de effectiviteit, de aanvaardbaarheid en de praktische haalbaarheid van gedragsinterventies, en meer specifiek nudges, in beleid te analyseren. (De Ridder et al., 2020). Als synthese-oefening werden drie conclusies naar voren geschoven. Ten eerste, nudges zijn in de meeste gevallen aanvaardbaar en vormen geen grote bereiding voor de keuzevrijheid en autonomie van burgers (Vugts et al., 2020). Ten tweede, nudges bieden geen garantie op succesvolle gedragsverandering, omdat de effectiviteit van nudges deels bepaald wordt door de individuele voorkeuren en "nudgeability" van burgers (de Ridder et al., 2021) . Ten derde, het is niet zo eenvoudig om deze simpele nudges zomaar in te passen in de complexe, weerbarstige en politiek gekleurde wereld waar het beleid tot stand komt (Feitsma, 2019). Deze drie conclusies bieden een mooie opstap naar en vergelijkingsbasis met de ambities van dit onderzoeksrapport

De structuur van het rapport ziet er als volgt uit. We starten met een algemene beschouwing op wat we het gedragskundige beleidsperspectief noemen (Hoofdstuk 1). Meteen daarna gaan we dieper in op de essentie van experimentele designs. We bekijken de belangrijkste ontwikkelingen, soorten, voorwaarden, troeven en grenzen van experimenten in bestuur en beleid (Hoofdstuk 2). Vervolgens brengen we verslag uit van twee experimenten die wij zelf in samenwerking met overheidsorganisaties hebben ontworpen, geïmplementeerd en geëvalueerd. Het eerste onderzoeksproject is een survey-experiment waarin we samen met de Vlaamse Stichting Verkeerskunde (VSV) het effect testen van rationele en emotionele boodschappen, toegevoegd aan de boete voor een snelheidsovertreding die in het veld wordt gebruikt (Hoofdstuk 3). Het

tweede onderzoeksproject betreft een veldexperiment waarin we samen met de dienst GASAM van de Stad Mechelen foutparkeerders via verschillende boodschappen hebben genudged om hun GAS-boetes sneller te betalen (Hoofdstuk 4). In het besluit bouwen we verder op de literatuur en op onze eigen praktijkervaring om te reflecteren op de belangrijkste praktische, methodologische en ethische aspecten bij het ontwerpen, opzetten, implementeren en evalueren van beleidsexperimenten op basis van gedragsinzichten.

1 HET GEDRAGSKUNDIG PERSPECTIEF

De gedragswetenschappelijke omslag of *behavioural turn* draait in de eerste plaats rond het integreren van psychologische inzichten en gedragskennis bij het bestuderen en sturen van individuele keuzes van burgers, ambtenaren en politici. Binnen de academische bestuurskunde vertaalt deze beweging zich in de opkomst van nieuwe sub-disciplines en tijdschriften zoals *Behavioural Public Policy* (Ewert, 2020; John, 2018) of *Behavioural Public Administration* (Battaglio et al., 2019; S. Grimmelikhuijsen et al., 2017). Binnen de beleidspraktijk uit dit zich in het uitvoeren van gedragsanalyses en het ontwerpen van gedragsinterventies. Het gedragsperspectief belooft een verrijking en een versterking van zowel beleidsonderzoek als beleidspraktijk, net zoals gedragseconomie een aanwinst is gebleken voor de neoklassieke economie. Gecharmeerd door het aantrekkelijke nudge-narratief van Thaler & Sunstein (2008) gingen overheden het potentieel van gedragsinzichten in snel tempo ontdekken en ontsluiten. In dit hoofdstuk situeren en presenteren we het gedragskundig perspectief ten opzichte van de bestaande perspectieven op het beleidsproces. Vervolgens belichten we waarom sommigen nudging als een zegen en anderen nudging als een vloek beschouwen voor de toepassing van gedragsinzichten in beleid. Tot slot trachten we los te komen van zowel het narratief als het kritisch perspectief door een functioneel perspectief te hanteren. Wat zijn nudges precies? Welke soorten zijn er? En welke vijf stappen moet je doorlopen om een gedragsinterventie of nudge op te zetten?

1.1 EEN VIJFDE BELEIDSPERSPECTIEF?

Binnen de beleidswetenschappen circuleren verschillende scholen, stromingen, paradigma'sen perspectieven (Howlett et al., 2009; Knoepfel et al., 2011). Bekkers et al. (2017) onderscheiden vier perspectieven op het beleidsproces. Zo zou het beleidsproces benaderd kunnen worden vanuit een rationeel, een politiek, een cultureel en een institutioneel perspectief. De onderverdeling van deze perspectieven is gebaseerd op zeven eigenschappen: het verklaringsmechanisme, het mensbeeld, de beleidsinstrumenten, de evaluatiecriteria, de rol van kennis en informatie, de structuur van het beleidsproces en de rol van politiek.

Het **rationele perspectief** kunnen we beschouwen als het referentieperspectief. De andere benaderingen vergelijken zich met, haken in op, of verzetten zich tegen het rationeel perspectief. Beleid wordt gezien als een middel om maatschappelijke problemen op te lossen en draait om het realiseren van doelstellingen.

Het belangrijkste verklaringsmechanisme is doelrationaliteit. Om die doelen te bereiken wordt gerekend op de causale keten tussen beleidsactie en beleidsresultaat. Effectiviteit en efficiëntie vormen de belangrijkste evaluatiecriteria. Wetenschappelijke kennis en objectieve informatie voeden het beleidsproces en moeten de keuze van beleidsinstrumenten onderbouwen. In de toolbox van de beleidsmakers zitten juridische, economische en communicatieve instrumenten, vaak omgedoopt tot stokken, wortels en preken (Vedung, 1998). Het dominante mensbeeld is dat van de *homo economicus*. Beleidsmakers en burgers opereren als rationale beslissers die de kosten en baten van verschillende (beleids)opties tegen elkaar afwegen (Leong & Howlett, 2020). Het beleidsproces is lineair of cyclisch en kan gevat en gepland worden volgens logisch op elkaar volgende stadia of fasen. De grens tussen politiek en administratie is helder en scherp. Het rationeel perspectief sluit aan bij de positivistische stroming in de sociale wetenschappen en baseert het zich op een werkelijkheid die meetbaar en kenbaar is.

Het **politieke perspectief** vertrekt vanuit de permanente machtsstrijd tussen verschillende actoren in het beleidsproces die vooral bezig zijn met het verdedigen en het promoten van waarden en belangen. De (vaak tegengestelde) waarden van deze actoren kunnen niet altijd op een rationele manier ten opzichte van elkaar uitgebalanceerd worden. Het beleidsproces lijkt op een arena waarin de verschillende spelers strijd met elkaar leveren via het voeren van onderhandelingen, afsluiten van compromissen, of aangaan van coalities. Ook drukking en dwang zijn niet vreemd aan de strijd om het beleid. Politieke en strategisch vaardigheden moeten helpen om de waarden en belangen veilig te stellen. Kennis, informatie en ook de beleidsinstrumenten van de overheid zijn niet langer rationele en neutrale bouwstenen, maar worden strategische en waarden-geladen radars of zelfs 'wapens' in het beleidsproces. De scheidslijn tussen politiek en ambtenarij is flou en doorwaadbaar. Het beleidsproces wordt gekenmerkt door dynamiek of zelfs chaos, en is meer incrementeel en iteratief. Het zet zich duidelijk af tegen het zeer instrumenteel-rationele perspectief (Bekkers et al., 2017).

Het **culturele perspectief** gaat ervan uit dat de wereld en de manier waarop wij die beschouwen sociaal geconstrueerd is. De wereld is niet objectief kenbaar en meetbaar, maar wordt gekneed door mensen en door de uiteenlopende manieren waarop zij met elkaar interageren en communiceren. De verschillende actoren in het beleidsproces trachten tot een gemeenschappelijk begrip van de werkelijkheid te komen door gebruik te maken van taal, symbolen, metaforen en afbeeldingen. De mens wil betekenis en zin geven door middel van gemeenschappelijke kaders en verhalen. Informatie en kennis kan geframed worden. Niet

zozeer effectiviteit, efficiëntie en causaliteit zijn belangrijke evaluatiecriteria, maar aspecten zoals kwaliteitsvolle argumentatie en open interactie staan centraal. Het beleidsproces is narratief en omvat een gemeenschappelijk discours dat continu geproduceerd en gereproduceerd wordt. Het (politieke) besluitvormingsproces vormt geen muur of machtsspel, maar een gemeenschappelijke onderneming om betekenis en zin te geven aan beleid. Het cultureel perspectief sluit aan bij de stroming van het constructivisme binnen de sociale wetenschappen en gaat uit van een werkelijkheid die sociaal geconstrueerd is en geïnterpreteerd wordt (Bekkers et al., 2017).

Het **institutionele perspectief** focust op de formele en informele regels, zoals procedures en routines, die het gedrag van de beleidsactoren beïnvloeden. De richting van beleid zou in grote mate worden bepaald door de context en de instellingen waarin het vorm krijgt. Dit perspectief laat weinig bewegings- en interpretatievrijheid aan individuele actoren. Het beleid is niet het resultaat van ratio, politieke macht of een welgekozen frame, maar van de sociale en historische inbedding ervan. De 'padafhankelijkheid' van beleid is een fundamenteel verklaringsmechanisme voor beleidskeuzes. Wanneer beleid al geruime tijd hetzelfde pad bewandelt, wordt het steeds moeilijker om op zijn stappen terug te keren of van pad te veranderen. Het mens- en wereldbeeld van het institutioneel perspectief wordt begrensd door formele en informele regels. Scherp gesteld, worden instellingen beschouwd als een soort ijzeren kooi, waarin de rol en de actieradius van de actoren op voorhand in grote lijnen vastliggen. De institutionele setting bepaalt welke informatie, welke beleidsinterventies of –instrumenten in overweging zullen worden genomen door beleidsactoren en welke methoden in aanmerking komen om interventies te testen. Ook het politieke speelveld is gebonden door deze institutionele factoren (Bekkers et al., 2017).

Deze verschillende perspectieven zijn conceptuele 'ideaaltypes, die daarom niet één op één terug te vinden zijn in de beleidspraktijk. Hun ideaalbeelden kunnen we zien als heuristieken die de beleidswerkelijkheid vereenvoudigen en dus deels ook reduceren (Wayenberg, 2020). De realiteit laat zich immers niet zo gemakkelijk capteren door één enkel perspectief. Vaak zullen meerdere perspectieven elkaar overlappen of elkaar aanvullen en zijn ze onderling combineerbaar. Bekkers et al. (2017) wijzen bijvoorbeeld zelf al op de raakvlakken tussen het rationele en het politieke perspectief enerzijds en tussen het culturele en institutionele perspectief anderzijds. De vier perspectieven zijn natuurlijk niet de enige poging om orde te scheppen in het beleidswetenschappelijke spectrum. De acht beleidsmodellen van Dye (2017), met o.a. het rationele keuzemodel en het institutionele model, volgt een gelijkaardige onderverdeling en logica.

Beleidsevaluatie kent ook tendensen of typen die tot op zekere hoogte dezelfde schikking voorstellen: rationeel-analytische, constructivistische en contextueel-realistische beleidsevaluatie (Van der Knaap et al., 2020).

De vier perspectieven bieden volgens Wayenberg (2020) een robuust raamwerk om de opkomst en de essentie van een vijfde perspectief te verklaren: het gedragskundige of behavioural perspectief (zie Tabel 1). Het verklaringsmechanisme voor het gedragskundige perspectief is de "begrensde rationaliteit", een begrip dat meer dan een halve eeuw reeds gemunt werd door Nobelprijswinnaar Herbert Simon (Olsen, 2015; Simon, 1955). Het concept wijst op de grenzen van het rationale denken over beleid. We kunnen er niet vanuit gaan dat politici, ambtenaren en burgers, overal en altijd volledig rationele beslissingen nemen. Het mensbeeld van het gedragskundige perspectief gaat uit van de feilbare *homo psychologicus* die beleid maakt op basis heuristieken (vuistregels) en vatbaar is voor cognitieve biases (Kahneman, 2012). De intellectuele wortels van het gedragskundige perspectief worden toegeschreven aan een diverse en toonaangevende groep gedragswetenschappers (waaronder enkele Nobelprijswinnaars) zoals Daniel Kahneman, Amos Tversky, Richard Thaler, Cass Sunstein en Robert Cialdini (Van Ryzin, 2021).

We beargumenteren in dit onderzoeksrapport dat het vijfde, gedragskundige perspectief het beleidsproces kan verrijken en versterken op twee manieren. Ten eerste zijn gedragswetenschappelijke inzichten reeds doorgedrongen tot de literatuur en het ontwerp van beleidsinstrumenten. Daar wordt het terrein grotendeels ingepalmd door nudges. Deze simpele en subtiele gedragsinterventies maken gebruik van keuze-architectuur om het gewenste gedrag zo gemakkelijk, aantrekkelijk, sociaal en tijdsgebonden mogelijk te maken - zonder zaken te verbieden of financieel te stimuleren/ontraden (BIT, 2014; Thaler & Sunstein, 2018). Er zijn ook nieuwere gedragsgeïnspireerde beleidsinterventies. Boosts willen het maken van gerichte en vooral bewuste keuzes versterken (Grüne-Yanoff & Hertwig, 2016; Hertwig & Grüne-Yanoff, 2017). Daarnaast is er sprake van een nudge plus beweging die deliberatieve en reflectieve elementen toevoegt aan de meer klassieke nudges zoals het gebruik van sociale normen of defaults (Banerjee & John, 2021). Ten tweede bieden gedragswetenschappelijke methoden extra mogelijkheden voor het testen van beleid. Hoewel het gedragskundig perspectief de meeste aandacht schenkt aan minder rationale en suboptimale denkprocessen, wordt eveneens opvallend veel belang gehecht aan een strakke evaluatiemethode op basis van experimentele designs. Die moet uitsluitsel bieden over de effectiviteit van gedragsinterventies en de causale mechanismen die daarachter schuilgaan (John, 2020). De methodologische aanpak uit

gedragswetenschappelijk onderzoek sluit aan bij de filosofie van *evidene-based policy making* en een variant hierop: het radicaal incrementalisme (Halpern & Mason, 2015). Er wordt bij voorkeur gewerkt met *randomized controlled trials* in het veld bij relatief kleine steekproeven om ex ante en stap voor stap te testen wat werkt en wat niet werkt. Indien de resultaten succesvol zijn, kunnen ze uitgerold worden voor de volledige populatie (Haynes et al., 2012).

Ondanks de toegevoegde waarde van het gedragskundige beleidsperspectief met betrekking tot nieuwe inzichten en methoden, zijn drie belangrijke nuances op hun plaats. Ten eerste moeten we het beeld van de *homo psychologicus*, die als een cognitieve kreupele de irrationele beslissingen en denkfouten opstapelt, bijstellen. Binnen de gedragswetenschappen bestaat er geen consensus over hoe stabiel en pertinent al die heuristieken en cognitieve biases wel zijn (Spiliopoulos & Hertwig, 2020). Voorstanders van de zogenaamde 'ecologische rationaliteit' stellen bijvoorbeeld dat heuristieken niet per definitie leiden tot denkfouten, maar net bijzonder nuttig zijn bij het maken van de meeste optimale keuzes (Gigerenzer et al., 2011). Ten tweede moeten we waarschuwen voor het risico op 'methodolotry'. Een te grote focus op één bepaalde methode, in dit geval het experimentele design, verhoogt de kans dat er enkel nog onderzoeks- en beleidsprojecten worden uitgevoerd die zich lenen tot het opzetten van experimenten (Moynihan, 2018). Hier komen we nog uitgebreid op terug in het volgende hoofdstuk over de experimentele methode. Ten derde moeten we opmerken dat binnen het gedragskundige perspectief op een bijzonder rationele manier over het beleidsproces wordt gedacht. Beleidsmakers kiezen op basis van een uitgebreide gedragsanalyse voor de geschikte gedragsinterventie, waarvan de effectiviteit rigoureus getest en geëvalueerd wordt (Feitsma, 2020). Toch zijn het beleidsproces zelf en zijn actoren niet onfeilbaar en worden beleidsmakers door dezelfde soort biases en heuristieken beïnvloed die ze beleidsmatig inzetten om burgers te beïnvloeden (Battaglio et al., 2019; S. Grimmelikhuijsen et al., 2017; Strassheim, 2019). Lodge & Wegrich (2016) noemden dit de 'rationaliteitsparadox'.

	Rational	Political	Cultural	Institutional	Behavioural
Verklarings -mechanisme	Doelrationaliteit	Macht en belangen	Zingeving en interactie	(in)formele regels, padafhankelijkheid	Begrensde rationaliteit
Menselijk gedrag	Homo economicus Kiest op basis van kosten en baten	Homo politicus Ziet de werkelijkheid als een politieke strijd om macht	Homo narratus Construeert de realiteit via frames en verhalen	Homo captus Zit gevangen in een institutionele kooi van regels en gedragingen	Homo psychologicus Laat keuzes onbewust beïnvloeden door heuristieken en biases
Beleids-instrumenten	Rationele selectie van beleidsinstrumenten uit de toolbox	Beleidinstrumenten als bron van macht	Beleidsinstrumenten faciliteren open, machtsvrije dialoog	Beleidsinstrumenten zijn institutioneel ingebed	Nudges, boosts en keuze-architectuur spelen in op onbewust gedrag
Evaluatiecriteria	Effectiviteit, efficiëntie en coherentie	Steun en draagvlak	Kwaliteit van argumentatie en discours	Logic of consequences Logic of appropriateness	Causaliteit en effectiviteit op basis van experimentele designs
Rol van kennis en informatie	Objectieve kennis en informatie over rationele keuzes van actoren leidt tot beter beleid	Kennis en informatie zijn bronnen van macht die waarden en belangen helpen verdedigen	Kennis wordt sociaal en cultureel geconstrueerd via frames en verhalen	Relevante kennis is afgebakend door de institutionele setting	Gedragswetenschappelijke inzichten over onbewust gedrag leidt tot beter beleid
Structuur van beleidsproces	Cyclisch en lineair met logisch op elkaar volgende fasen	Grillig, dynamisch en incrementeel	Uitwisselen van ideeën, ervaringen en observaties richting een gemeenschappelijk discours	Bepaald door informele en formele regels	Rationeel incrementalisme
Rol van politiek	Duidelijk scheiding tussen politiek en administratie	Politiek als machtsspel	Politiek als zingeving, die ruimte laat voor dialoog en interactie	Politiek volgt de institutionele logica	Politiek is evenmin immuun voor cognitieve biases

Tabel 1. Vijf perspectieven op het beleidsproces (Bekkers et al., 2017; Wayenberg, 2020)

1.2 NUDGING IN PERSPECTIEF

Het fenomeen nudging kunnen we eveneens vanuit verschillende perspectieven benaderen. We onderscheiden drie belangrijke invalshoeken: het narratieve perspectief, het kritische perspectief en het functionele perspectief (Raymaekers & Brans, 2020).

Ten eerste wordt **het narratieve perspectief** gehanteerd door voorstanders van nudging die het zien als een zegen. Nudging geeft een realistisch in plaats van een economisch mensbeeld, wat wordt uitgerukt in het verschil tussen "*econs*" en "*humans*" (Thaler, 2015). Nudging bouwt voort op de kennis en inzichten die gedurende decennia werden opgebouwd binnen de gedragswetenschappen. De basisconstructen zijn de *dual process theory* (met het snelle, automatische systeem-I denken en het trage, reflectieve systeem-2 denken) alsook een uitgebreide catalogus met heuristieken en cognitieve biases (Kahneman et al., 1982). Heuristieken en biases maken het gedrag van mensen niet alleen feilbaar, maar tot op zekere hoogte ook voorspelbaar en op die manier kneedbaar. Nudges worden gepresenteerd als eenvoudige en tegelijk slimme aanpassingen die weinig kosten, de keuzevrijheid en autonomie van de individuele burger vrijwaren, vertrekken vanuit het 'echte' gedrag van mensen en via een wetenschappelijke toetsing kunnen leiden tot relatief grote resultaten (Benartzi et al., 2017; Thaler & Sunstein, 2018). Het nudgen zelf gebeurt vaak via subtiele aanpassingen aan geijkte, administratieve procedures zoals het versturen van brieven, mails, formulieren of folders (Hallsworth & Kirkman, 2020)en wordt toevertrouwd aan gedragsexperts die deel uitmaken van gedragsteams en nudge-netwerken, en opereren binnen en doorheen de verschillende overheidsdepartementen (Feitsma, 2019a; John & Stoker, 2019).

Ten tweede is er **het kritische perspectief,** waarbij nudging niet als een zegen, maar als een vloek wordt gepercipieerd. Nudging mag dan populair zijn, daarom is het nog niet vrij van controverse. Er woedt een hevig academisch debat over de valkuilen en zelfs 'the dark side" van nudging (Madi, 2020). De kritiek valt uiteen in drie grote dimensies (Raymaekers et al., 2019). Tegenstanders bekritiseren nudging op basis van epistemologische grenzen, ethische risico's en ideologische bijbedoelingen. Nudging wordt soms weggezet als een overschatte hype die nauwelijks substantiële beleidsvoordelen oplevert (Whitehead et al., 2020). Een belangrijke bron van kritiek richt zich op de beperkte effectiviteit van nudges. Werken ze wel en hoe lang

werken ze door? Verschillende reviews en meta-analyses komen met gemengde resultaten. In sommige domeinen, zoals het stimuleren van gezonde voeding (Arno & Thomas, 2016), zien we behoorlijke grote effecten, terwijl nudging in een belastingcontext dan weer erg beperkte resultaten oplevert (Antinyan & Asatryan, 2019). Een generieke meta-analyse van Hummel & Maedche (2019) op basis van 100 papers met in totaal 31 effectmetingen van nudges kwam uit bij een mediaan relatief effect van 21% en een gemiddeld relatief effect van 55%. We weten echter dat *publication bias*, namelijk het niet of minder publiceren van nudge-experimenten met nulresultaten, een belangrijke factor is (Della Vigna & Linos 2020). Ook over de langetermijneffecten van nudges tasten we nog grotendeels in het duister (Allcott & Rogers, 2014). Verder is er kritiek op de begrensde epistemologie (uitsluitende focus op het individuele keuzeproces) en de eenzijdige methodologie (het gebruik van enkel RCT's) van nudges (MacKay & Quigley, 2018; Madi, 2020; Mols et al., 2015). Daarnaast schuiven critici ethische bezwaren naar voren. Ze schetsen het beeld van beleidsmakers als stiekeme keuze-architecten die door middel van psychologische trucjes en manipulatieve technieken bewust inspelen op het onbewuste keuzegedrag van burgers (Goodwin, 2012; Wilkinson, 2013). Aansluitend wordt gewezen op een gebrek aan zichtbaarheid en transparantie (Hansen & Jespersen, 2013; Lepenies & Malecka, 2019). Tot slot speelt ook ideologie een rol. Hoewel nudging volgens Sunstein & Thaler (2003) voortbouwt op de politieke filosofie van het libertair paternalisme, kan het gebruik van nudging voor de ene criticaster overkomen als té paternalistisch (betuttelend en invasief) en voor de ander dan weer als té libertair (een vrijblijvende afleiding van echte beleidsproblemen en –oplossingen) (Jones et al., 2013) .

Ons interesseert in het kader van dit onderzoeksrapport over *nudging in actie* vooral een derde perspectief. **Het functionele perspectief** gaat over hoe beleidsmakers concreet aan de slag moeten met nudging en de bredere beweging van de gedragsinzichten.

1.3 HET FUNCTIONELE PERSPECTIEF

In ons vorige rapport hadden we de brede familie van nudges al onderverdeeld in vier overkoepelende typen. Er zijn bovendien twee benaderingen voor het toepassen van nudges, en verschillende frameworks die beleidsmakers kunnen hanteren. We lichten de verschillende elementen van het functionele perspectief hieronder verder toe.

De vier types nudges zijn 1) het aanpassen van de informatie-omgeving door vereenvoudiging of framing; 2) het aanpassen van de fysieke omgeving; 3) het aanpassen van de sociale omgeving; en 4) het aanpassen van de standaardoptie of default (House of Lords Science and Technology Select Comittee, 2011; Lehner et al., 2016). Klassieke voorbeelden van deze vier types nudges zijn: het visualiseren van informatie via bijvoorbeeld het Europese energielabel of de Nutri-score; het plaatsen van fruit op ooghoogte; het toevoegen van een sociale norm aan de belastingbrief; of groene energie als het standaardpakket aanbieden. Recent kunnen we daaraan toevoegen het tekenen van krijtcirkels op grote pleinen om de coronabubbels onder controle te houden (Raymaekers, 2020; Raymaekers & Migchelbrink, 2021).

In grote lijnen kunnen we twee verschillende benaderingen voor de toepassing van nudges onderscheiden. Een eerste visie vertrekt vanuit de strikte definitie van nudges en beschouwt ze als een afzonderlijk, vierde type beleidsinstrument (de Ridder & Tummers, 2019; Tummers, 2019a). Nudges komen in dat geval naast de wortel, de stok en de preek, en kunnen dan ook apart van deze traditionele instrumenten worden ingezet. Een tweede visie ziet nudges minder als een afzonderlijk beleidsinstrument en meer als technieken om de traditionele instrumenten inventiever en effectiever te maken. Nudges doen in dit geval dienst als een soort hefboom, ze voegen een zogenaamde 'behavioural spin' toe aan traditionele instrumenten (Loer, 2019). De beleidsmaker kijkt dan door een 'gedragslens' naar het volledige beleidsplaatje om de effectiviteit en efficiëntie van zijn beleidsinstrumentarium te verhogen (Ewert, 2020).

Om de overstap van theorie naar praktijk te vergemakkelijken heeft de beleidsmaker ondertussen ook een bijzonder breed gamma aan functionele frameworks ter beschikking -met MINDSPACE en EAST als bekendste exemplaren. Daarnaast zijn er verschillende draaiboeken, stappenplannen, kompassen, gedragstoetsen, gedragsverkenners en zelfs gedragswielen ontwikkeld (voor een uitgebreid overzicht van kaders en tools, zie Raymaekers et al., 2019). In grote lijnen moeten bij het opzetten van een gedragsinterventie steeds dezelfde vijf fasen worden doorlopen.

1. De definitiefase: In de eerste fase moeten we zowel het probleemgedrag als het gewenste gedrag (ook doelgedrag genoemd) zo scherp mogelijk definiëren. Wat is het probleemgedrag? Wie is

de doelgroep? Wie doet wat niet goed, waar en wanneer? En waarom is dat juist een probleem? Deze zogenaamde vijf W's kunnen ons op weg helpen.

2. **De analysefase:** De tweede fase bevat een gedragsanalyse waarbij de gedragsbepalende factoren of determinanten van het gedrag worden onderzocht. Er bestaat een hele reeks gedragsmodellen en – theorieën die als vertrekbasis kunnen dienen (Kok et al., 2016). Het COM-B systeem biedt bijvoorbeeld een eenvoudig model om de drivers en barrières van gedrag te begrijpen. Gedrag (Behaviour) wordt bepaald door de interactie tussen drie componenten: Capaciteit, Opportuniteit en Motivatie (Michie et al., 2011). Capaciteit kijkt naar de psychische en fysieke mogelijkheden van mensen, naar de vereiste kennis en vaardigheden om het gewenste gedrag te stellen. Motivatie gaat over de automatische en reflectieve cognitieve processen die ons gedrag aansturen. Motivatie bevat niet alleen het doordachte, analytische beslissingsproces, maar ook onze routineuze, emotionele en instinctieve reacties. Opportuniteit heeft betrekking op alle factoren die zich buiten het individu bevinden, zoals de fysieke en de sociale omgeving. Die maken het gewenste gedrag mogelijk of lokken het zelfs uit. Om barrières te detecteren en gedrag te begrijpen is het belangrijk om op basis van deze 3 hefbomen de juiste vragen te stellen.

3. **De interventiefase:** In de derde fase inventariseren en selecteren we de gedragsinterventie(s). Dit kan gebeuren op verschillende manieren. Zo kunnen we werken op basis een grondige literatuurstudie naar de effectiviteit en repliceerbaarheid van verschillende interventies. Daarnaast kunnen ook bepaalde frameworks soelaas bieden. Het meest gebruiksvriendelijke en allicht daarom ook populairste model is EAST. Het werd in 2014 ontworpen door het Britse *Behavioural Insights Team* (BIT, 2014; Halpern, 2015). De basisgedachte van EAST is simpel: als je een bepaald gedrag wil aanmoedigen, kan je het best zo **E**asy, **A**ttractive, **S**ocial, **T**imely, en dus zo "EAST" mogelijk maken.

4. **De designfase:** In de vierde fase gaan we nadenken over het ontwerp en de implementatie van de interventie. Het *Behavioural Insights Netwerk* Nederland heeft hiertoe een handig overzicht

ontwikkeld met 18 onderzoeksmethodieken (BIN NL, 2019). Dit overzicht bevat kwalitatieve methoden zoals diepte-interviews, focusgroepen en dagboekmethoden, maar ook meer kwantitatieve methoden zoals surveys, RCT's, regressie-analyses en difference-in-difference designs.

5. **De evaluatiefase:** De laatste fase omvat het evalueren van de gedragsinterventies. Door de data te analyseren, zullen we op basis van ons gekozen design kunnen vaststellen of en in welke mate de gedragsinterventies effectief zijn geweest en of ze het gedrag ook in de gewenste richting hebben kunnen beïnvloeden (Van der Knaap et al., 2020). Een kritische blik en de nodige bijsturing zijn hier cruciaal om beleidskeuzes stevig te onderbouwen.

In het volgende hoofdstuk zullen we dieper ingaan op één mogelijk design, namelijk de experimentele methode. Meer specifiek kijken we naar ontwikkelingen, de slaagkansen, de troeven en de valkuilen van beleidsexperimenten. In de twee daaropvolgende hoofstukken zullen we stilstaan bij twee beleidsexperimenten die we zelf hebben uitgevoerd.

2 DE EXPERIMENTELE METHODE

De experimentele methode wint aan populariteit en autoriteit binnen de bestuurs- en beleidswetenschappen (Anderson & Edwards, 2015; Bouwman & Grimmelikhuijsen, 2016; Hansen & Tummers, 2020; Li & Van Ryzin, 2017). Deze trend kan deels geïnterpreteerd worden als een reactie op en een erkenning van de beperkingen van niet-experimentele observatiemethoden zoals vragenlijsten, interviews en focusgroepen (Baekgaard et al., 2015). Daarnaast is er de toegenomen appreciatie voor causaliteit, de oorzaak-gevolg relatie tussen twee variabelen. Een correlatie is immers niet hetzelfde als een causaal verband tussen twee variabelen. Een correlatie tussen het aantal bosbranden en de verkoop van ijsjes betekent bijvoorbeeld nog niet dat er causaal verband bestaat tussen beide fenomenen.

Zowel onderzoekers als beleidsmakers zijn over het algemeen geïnteresseerd in de attributievraag: is X de oorzaak van Y en wat is de effectgrootte van X op Y (James, John, et al., 2017). Als we dit naar het beleidsproces vertalen komt dit neer op het zogenaamde effectiviteitsvraagstuk van beleid, in hoeverre kan een beleidsresultaat rechtstreeks worden toegeschreven aan een bepaalde beleidsactie. Als we die lijn verder doortrekken naar het gedragsperspectief dan willen we weten in hoeverre een bepaalde gedragsinterventie (of nudge) leidt tot een gewenste gedragsverandering. Om die vraag te beantwoorden kan in sommige situaties een experimenteel design aangewezen zijn. Wie daarbij denkt dat experimenten een eenvoudig toe te passen mirakeloplossing bieden voor de limieten van observationele data, kan echter bedrogen uitkomen. Ook bij experimenten komen immers een hele reeks praktische, methodologische en ethische uitdagingen kijken (Baekgaard et al., 2015).

In dit hoofdstuk bieden we onderzoekers en practici de fundamentele bouwstenen voor het opzetten van experimenten in bestuur en beleid. Voor meer geavanceerde en exhaustieve kennis over de verschillende type experimenten, experimentele designs en statistische analysetechnieken verwijzen we door naar meer gespecialiseerde studies (Brown & Melamed, 1990; Gerber & Green, 2012; Kirk, 2012; Shadish et al., 2002). In wat volgt bieden we een begin van antwoord op vijf vragen. Wat zijn experimenten? Welke soorten experimenten bestaan er? Waarom experimenteren in bestuur en beleid? Hoe experimenteren in bestuur en beleid? En waar liggen de grenzen?

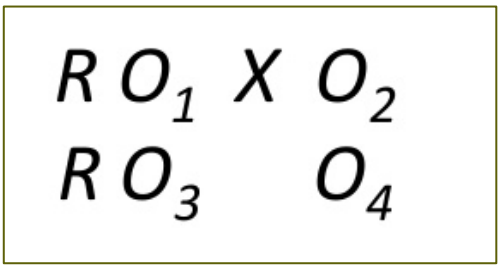
2.1 WAT IS EEN EXPERIMENT?

Een experiment is een geijkte, wetenschappelijke methode om kennis te verwerven door (theoretisch) veronderstelde causale verbanden te toetsen. Centraal staat het concept van de *Randomized Controlled Trial*. Om te spreken van een klassiek of zuiver experiment moeten drie voorwaarden vervuld zijn (van Thiel, 2015):

1. De proefpersonen worden wikkeurig en dus per toeval (at random = R) verdeeld in twee (of meer) groepen: een controlegroep en minstens één experimentele groep (of interventiegroep).

2. De interventiegroep ondergaat de interventie (X), de controlegroep niet.

3. Bij beide groepen wordt een voormeting (O_1) en een nameting gehouden (O_2), zodat het effect van X kan worden berekend.[1]

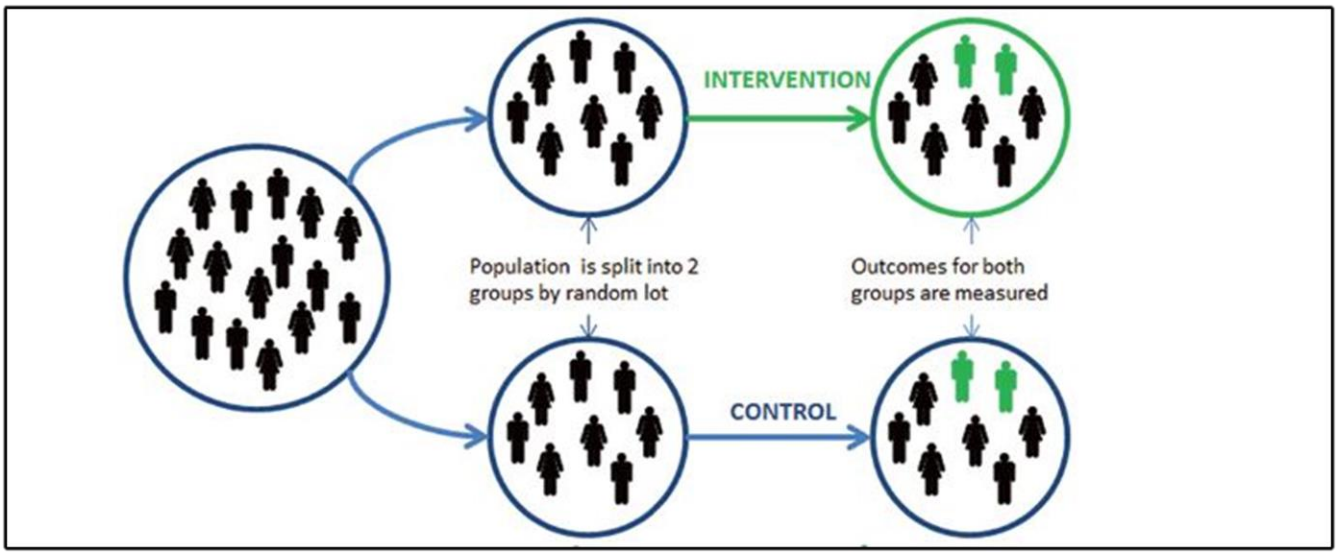


Figuur 1. Schematische voorstelling van een klassiek experiment met een pretest posttest control group design (van Thiel, 2015)

Op basis van het experimenteel design kan precies worden vastgesteld wat het verschil is tussen beide groepen en wat het effect is van de interventie of de onafhankelijke variabele op de te meten of afhankelijke variabele. Daarmee bieden experimenten een oplossing voor de hardnekkige onderzoeksproblemen inzake endogeniteit (Morton & Williams, 2010). Endogeniteit impliceert dat het effect van X op Y niet correct geïnterpreteerd kan worden omwille van verdoken oorzaken, die buiten de experimentele setting liggen

[1] We zullen verderop zien dat een experimenteel design niet altijd een voormeting vereist. Men kan ook enkel op basis van de nameting de uitkomst van de interventie met de controle vergelijken en op basis daarvan conclusies formuleren. Een voormeting kan echter wel steeds nuttig zijn om voor elke groep te weten wat de startpositie (baseline) is van de afhankelijke variabele die onderzocht en gemanipuleerd wordt.

(Antonakis et al., 2010) Endogeniteit verwijst o.a. naar het verband tussen een variabele die opgenomen is in het design en een 'verstorende' variabele die niet opgenomen is. Denk aan de buitentemperatuur in het voorbeeld van de bosbranden en de ijsjes. Endogeniteit kan ook voorvallen wanneer er onduidelijkheid is over de richting van het causale verband (Bouwman & Grimmelikhuijsen, 2016). Zorgen jobcoaches voor een hogere motivatie bij werkzoekenden, of zijn sterk gemotiveerde werkzoekenden meer geneigd om een jobcoach te raadplegen. Een experimenteel design verhelpt deze en andere problemen en is daarom uitermate geschikt om de oorzaak-gevolg relatie tussen twee variabelen vast te stellen. In Figuur 1 en 2 geven we de drie criteria van een klassiek experiment schematisch weer:



Figuur 2. Het basisprincipe van een Randomized Controlled Trial (Haynes et al., 2012, 4)

Er zijn verschillende experimentele designs denkbaar (Blom-Hansen et al., 2015). Het schema in Figuur 1 omvat een voor- en een nameting en komt overeen met het zogenaamde *pretest posttest control group design*. Wanneer we enkel een nameting voorzien bij zowel de controle- als de interventiegroep spreken we over een *posttest only control group design*. Om zicht te krijgen op langere termijn effecten kan de nameting verschillende keren herhaald worden, en kunnen we nagaan of het effect afzwakt over tijd. Op dezelfde manier is het ook mogelijk om een hele reeks observaties zowel voor als na de interventie uit te voeren. In dat geval spreken we van tijdreeksen. Een overzicht van deze vier basisdesigns is te vinden in Tabel 2. Het design kan verder uitgebreid worden met bijvoorbeeld meerdere interventies, verschillende

interventiedosissen en hun onderlinge interactie via een factoriale experimenteel design. Gedetailleerde informatie over deze en meer complexe en geavanceerde designs, meetmethoden en analysetechnieken vallen buiten de scope van deze basisinleiding, maar is te vinden in de uitgebreide literatuur over experimenten in de sociale wetenshappen (Brown & Melamed, 1990; Gerber & Green, 2012; Kirk, 2012; Shadish et al., 2002)

Tabel 2. Experimentele designs (James et al., 2017)

	posttest only	pretest - posttest	≠ posttesten	tijdreeksen
Interventie:	X O	O X O	X O . . . . . O	O O O X O O O
Controle:	O	O O	O . . . . . O	O O O O O O

Om te kunnen spreken van een zuiver experiment dient de randomisatie, het per toeval indelen van individuen in groepen, steeds op een systematische manier te gebeuren. Elke onderzoekseenheid of proefpersoon heeft een gelijke kans om ofwel in de interventiegroep ofwel in de controlegroep te belanden. Op die manier kan men equivalentie tussen de groepen tot stand brengen. Equivalentie betekent dat de leden van elke groep dezelfde relevante kenmerken hebben (identieke samenstelling), even vatbaar en open zijn voor de interventie (identieke instelling) en hetzelfde proces doorlopen (identieke ervaringen) (De Peuter et al., 2007). Op die manier zijn de relevante en niet relevante variabelen in gelijke mate in alle groepen aanwezig waardoor hun verstorende werking evenwichtig over alle groepen verdeeld is.

De randomisatie kan op verschillende manieren ingericht worden. Naast individuele randomisatie kan ook gebruik gemaakt worden van blokrandomisatie of clusterrandomisatie (Gerber & Green, 2012; James, Jilke, et al., 2017a). Bij **blokrandomisatie** worden de onderzoekseenheden verdeeld in kleinere subgroepen of blokken. Binnen elk blok gaat men dan individueel randomiseren zodat er een reeks van mini-experimenten plaatsvindt. Op deze manier vergroot men de kans dat de eigenschappen tussen beide interventiegroepen gelijk zijn. Bij **clusterrandomisatie** wordt de interventie toegediend op niveau van de groep of cluster in plaats van op individueel niveau. De randomisatie gebeurt dan per cluster, denk bijvoorbeeld aan klassen, scholen of organisaties. Soms kan de keuze voor een geclusterd design ook noodzakelijk zijn omdat het niet mogelijk is om onderzoekseenheden op individuele basis te verdelen (zie ook Hoofdstuk 4).

Wanneer één van de drie cruciale voorwaarden ontbreekt, kunnen we niet meer spreken van een zuiver experimenteel design, maar van een **quasi-experimenteel design** (Shadish et al., 2002). Dit is bijvoorbeeld het geval wanneer er geen systematische randomisatie mogelijk is of wanneer de controlegroep ontbreekt en er meerdere voor- en nametingen worden gehouden binnen één en dezelfde groep. Bij een quasi-experimenteel design verhoogt de kans op endogeniteit, en dus vertekende resultaten (bias) en verlaagt de interne validiteit waardoor het lastiger is om causale verbanden vast te stellen (Eden, 2017). Toch zullen we zien dat bij beleidsexperimenten in het veld, waar volledige controle over de experimentele setting allesbehalve evident is, onderzoekers en beleidsmakers soms noodgedwongen werken met quasi-experimenten (John, 2017c). Tot slot vermelden we dat er ook zoiets bestaat als een **natuurlijke experiment**. Bij een natuurlijk experiment krijgen de onderzoekseenheden een interventie toegediend, maar deze werd bedacht noch bepaald door de onderzoekers. Er is geen bewuste manipulatie en randomisatie, maar de basis van vergelijking tussen groepen is het gevolg een bepaalde beslissing of gebeurtenis die daarbuiten valt (Dunning, 2012; Shadish et al., 2002). Denk bijvoorbeeld aan een organisatorische verandering, zoals een fusie tussen gemeente en OCMW, of aan een abrupte gebeurtenis zoals een overstroming, waarbij de ene regio een beroep kan doen op een specifieke steunmaatregel en de andere niet.

2.2 WELK TYPE EXPERIMENTEN?

Klassieke experimenten worden doorgaans geassocieerd met een laboratoriumsetting en witte stofjassen. Het klopt inderdaad dat de onderzoeker het liefst zoveel mogelijk controle heeft over de verschillende componenten van het experiment. De variabelen die onderdeel zijn van de onderzoekssituatie, zoals de beleidsmaatregel en het te meten effect, worden **endogene variabelen** genoemd. De onderzoeker richt de testomgeving op zo'n manier in dat hij of zij weet dat de proefpersonen door niets anders dan de experimentele stimuli worden beïnvloed. De onderzoeker bepaalt de interventie of de 'manipulatie' die proefpersonen ondergaan en zorgt ervoor dat alle invloeden van buitenaf (de **exogene variabelen**) zoveel mogelijk worden buitengesloten (van Thiel, 2015). Een ander belangrijk onderscheid is het verschil tussen interne en externe validiteit. **Interne validiteit** omvat de mate waarin causale (oorzaak-gevolg) relaties tussen de onafhankelijke variabele (beleid) en de afhankelijke variabele (effecten) correct onderzocht zijn waardoor ze intern geldig zijn. **Externe validiteit** heeft betrekking op de mate waarin de onderzoeksresultaten te generaliseren zijn naar andere populaties, de mate waarin ze ook extern geldig zijn en dus veralgemeend

kunnen worden (Shadish et al., 2002; Tepe & Prokop, 2017). De verschillende type experimenten kunnen we van elkaar onderscheiden op basis van een continuüm tussen controle en *fieldness*. Hieronder bespreken we de drie belangrijke type experimenten: labo-experimenten, veldexperimenten en survey-experimenten. Daarbij maken we gedeeltelijk gebruik van het zogenaamde UTOS model (Shadish et al., 2002), door aandacht te besteden aan de onderzoekseenheden (Units), de interventies (Treatments), de Observaties en de Settings Tot slot geven we de belangrijkste kenmerken van elk type weer in Tabel 3..

Labo-experimenten bieden in principe de meest gecontroleerde setting voor het uitvoeren van een experiment. Of het experiment plaatsvindt in een standaard klaslokaal of in een hoogtechnologisch laboratorium, de onderzoeker ambieert steeds maximale controle over zowel de endogene als de exogene variabelen en streeft naar de ideale omstandigheden om causale relaties te ontwaren met een hoge mate van interne validiteit (Tepe & Prokop, 2017). Labo-experimenten krijgen ook kritiek te verduren omwille van hun beperkte generaliseerbaarheid en het gebrek aan realisme (Falk & Heckman, 2009; Levitt & List, 2007) Zo treden studenten wel vaker op als semi-vrijwillige onderzoekseenheden binnen een labostudie. Experimenten in het laboratorium zijn gangbaar binnen economie (bv. financiële incentives) en psychologie (bv. cognitieve testen), maar komen minder vaak voor binnen onderzoek naar bestuur en beleid. De meeste onderzoeksonderwerpen in bestuur en beleid zijn immers niet eenvoudig vatten en te vangen in een laboratorium (Margetts, 2011). Dus hoewel labo-experimenten bijzonder waardevolle inzichten kunnen opleveren, die complementair zijn aan de input en de uitkomsten van veldexperimenten (Falk & Heckman, 2009), is dit type experiment minder geschikt om de effectiviteit van beleid vast te stellen.

Veldexperimenten vinden plaats in de echte wereld. De term verraadt de historische invloed van landbouwonderzoek, waar de experimentele methode op basis van randomisatie en controlegroepen gebruikt werd om de groei van gewassen te bestuderen en te verbeteren (Fisher, 1926). Veldexperimenten hebben minder controle in vergelijking met labo-experimenten, maar scoren dan weer erg hoog op de factor realisme. De mate waarin experimenten in de echte wereld (het veld) plaatsvinden, de zogenaamde *fieldness*, wordt bepaald door minstens vier criteria: de interventie is realistisch; de context is natuurlijk; de deelnemers ondergaan de interventie in de echte wereld, en de uitkomsten zijn authentiek (Eden, 2017; Hansen & Tummers, 2020). Czibor et al. (2019) voegen daar nog enkele criteria aan toe: bestaat de steekproef

uit studenten of uit de doelgroep en zijn de interventies transparant of niet. Veldexperimenten zijn in trek omdat ze een hoog potentieel hebben op het vlak van interne én externe validiteit. Sommige onderzoekers menen echter dat veldexperimenten per definitie quasi-experimenten zijn omdat het onmogelijk is om altijd en overal de omgeving en de setting volledig te controleren (van Thiel, 2015). Bovendien wijkt ook het design vaak af van een zuiver experiment, bijvoorbeeld als het gaat om de selectie van proefpersonen, waardoor men aan interne validiteit moet inboeten. Veldexperimenten zijn doorgaans het resultaat van een samenwerking tussen academische onderzoekers en overheidsorganisaties en zijn bij uitstek geschikt om de effectiviteit van beleid vast te stellen.

Survey-experimenten zijn ingebed in een online of offline vragenlijst en bezitten een eigen dynamiek en positie. Ze ontbreken de controle van een labo-experiment en het realisme van een veldexperiment. Context en contact kunnen via een survey immers sterk verschillen van een 'real-life' situatie. Hierdoor is het niet altijd duidelijk hoe survey de zelf-gerapporteerde motivatie en intenties van survey respondenten zich zouden vertalen naar gedragingen in de echte wereld (Levitt & List, 2007). Een aantal uitdagingen zijn dezelfde als voor veld of labo-experimenten, maar daarnaast moeten survey-experimenten ook rekening houden met een reeks praktische kwesties die gepaard gaan met het opstellen van een vragenlijst. Denk aan de volgorde, de toon en de presentatie van de vragen en mogelijke antwoorden (Krosnick, 1999; Thau et al., 2021). Ook een (te) lage responsgraad kan een probleem vormen. Voor survey-experimenten bij burgers wordt daarom soms een beroep gedaan op marktonderzoeksbureaus. Tegenover deze aandachtspunten staan ook enkele, duidelijke pluspunten. Zo zijn survey-experimenten in vergelijking met andere typen relatief goedkoop, hebben ze potentieel een groot bereik en kan men binnen één enkele survey meerdere experimenten opzetten. Bovendien hebben de meeste online surveytools zoals Qualtrics gebruiksvriendelijke randomisatiemogelijkheden ingebouwd die het ontwerp en de implementatie van survey-experimenten faciliteren (Li & Van Ryzin, 2017). Dit type experimenten is vooral geschikt om veranderingen in opvattingen en attitudes te meten, niet om concrete gedragswijzigingen vast te stellen.

Tabel 3. Kenmerken per type experiment

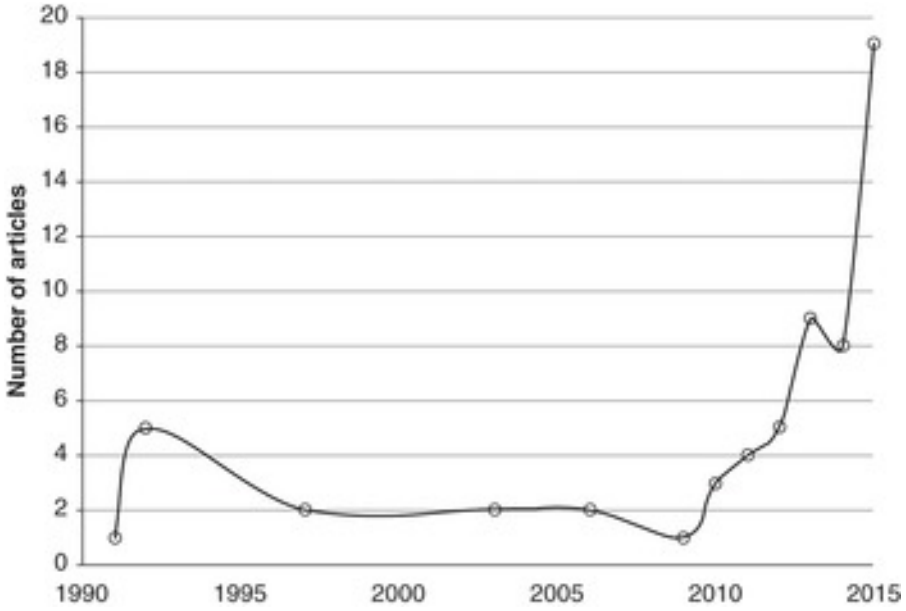
	Labo	Veld	Survey	Quasi
Manipulatie via OV	+	+	+	+
Gerandomiseerd	+	+	+	-
Gecontroleerd	+	-	+-	-
Interne validiteit	+	+-	+	-
Externe validiteit	+-	+-	+-	+-
Realisme	-	+	+-	+
Effectiviteit van beleid	-	+	+-	+-

(Geïnspireerd op Baekgaard et al., 2015; Hansen & Tummers, 2020; Jams et al., 2017; Jilke & Van Ryzin, 2017; John, 2017b; Podsakoff & Podsakoff, 2019; Tepe & Prokop, 2017)

2.3 WAAROM EXPERMINENTEREN IN BESTUUR EN BELEID?

Een aantal wetenschappelijke disciplines hebben een rijke methodologische traditie op het vlak van experimenten. Denk aan de gedragswetenschappen (Field & Hole, 2002) of de medische wetenschappen (Friedman et al., 2010). Inzichten en lessen uit studies naar de fysieke en mentale gezondheid van proefpersonen kunnen nuttig zijn voor bestuurskundigen en beleidswetenschappers, maar bieden niet altijd antwoorden op specifieke bestuurskundige onderzoeksvragen, die van toepassing zijn op een andere onderzoekscontext (de overheid) en met andere onderzoeksobjecten (ambtenaren, beleidsmakers en burgers) (James, John, et al., 2017; John, 2017c). De huidige interesse voor experimentele designs in onderzoek naar bestuur en beleid wordt aangewakkerd door het toenemende gebruik van deze methode in belendende disciplines zoals sociale wetenschappen (Druckman et al., 2011), politieke wetenschappen (Morton & Williams, 2010) en economische wetenschappen (Frechette & Schotter, 2015; Friedman, 2010; Harrison & List, 2004; Nagatsu & Favereau, 2020). Ook in managementstudies en onderzoek naar leiderschap winnen experimentele methoden terrein (Podsakoff & Podsakoff, 2019). Recent zien we ook een enorme toename van zogenaamde business experimenten (Thomke, 2020), waarbij grote bedrijven zoals Google, Amazon, Booking.com en Uber per jaar duizenden kleine A/B testjes inrichten om te achterhalen wat aanslaat bij hun consumenten en zo het ontwerp en de verkoop van hun producten en diensten te verbeteren (Luca & Bazerman, 2020; Van Ryzin, 2021).

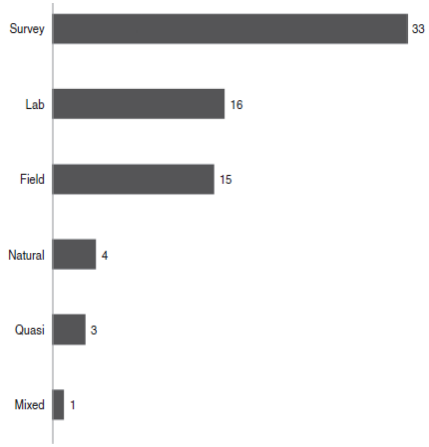
De aandacht voor experimenten binnen onderzoek naar bestuur en beleid was lange tijd bijzonder beperkt (Hansen & Tummers, 2020; Margetts, 2011). Volgens Bozeman (1992, pp. 440–441) kwam dit omdat bestuurs- en beleidswetenschappers meer interesse hebben voor het probleem dan voor de probleemoplossende technieken. Het voorbije decennium zien we op het onderzoeksterrein een duidelijke kentering (Bouwman & Grimmelikhuijsen, 2016). Op basis van een systematische review identificeerden Li & Van Ryzin (2017) 72 experimenten in internationale bestuurs- en beleidskundige tijdschriften tussen 1990 en 2015, met een opvallende versnelling in de laatste jaren.



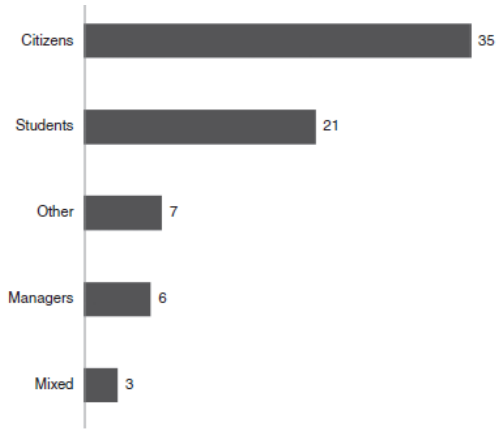
Figuur 3. Aantal gepubliceerde experimenten in de top 20 tijdschriften over bestuurs- en beleidskunde 1990-2015 (Li & Van Ryzin, 2017, p. 23)

Verder blijkt uit Figuur 4 dat vooral de survey-experimenten (33) erg in trek zijn met bijna de helft van het totale aantal experimenten, gevolg door de labo-experimenten (16) en de veldexperimenten (15). Daarnaast kunnen we kijken naar de onderzoekseenheden. Figuur 5 toont dat de meeste experimenten burgers als onderzoekseenheden nemen. Dit is vooral het geval bij survey-experimenten (25/33). Verder wordt gekeken in de richting van studenten, die gerekruteerd worden binnen de sociale wetenschappen. Zo stellen we vast dat alle labo-experimenten gebruik maken van studenten (16/16) (Li & Van Ryzin, 2017). Het valt ook op dat publieke managers zelf maar beperkt opduiken in een experimentele setting. Als we kijken naar specifieke thema's zien we dat burgerparticipatie (11) erboven uitsteekt gevolg door experimenten over prestatie-

informatie (9), het beslissingsproces (7) en motivatie (7). Over het algemeen merken we een tendens om bij experimenten te focussen op het individuele gedrag van burgers en in mindere mate publieke managers en beleidsmakers.



Figuur 4. Type experimenten (Li & Van Ryzin, 2017, p. 26)



Figuur 5. Type onderzoekseenheden (Li & Van Ryzin, 2017, p. 28)

Experimenten werden o.a. ingezet om de causale relatie te onderzoeken tussen de representativiteit van overheidsdiensten en de gepercipieerde legitimiteit ervan door burgers (Riccucci et al., 2014); tussen de transparantie van overheidsorganisaties en de gepercipieerde legitimiteit ervan door burgers (Grimmelikhuijsen & Meijer, 2014); tussen de representativiteit van overheidsdiensten en de bereidheid van burgers om te coproduceren (Riccucci et al., 2016; Sievert, 2021); tussen de opkomst en representativiteit van burgers bij participatieprojecten en de bereidheid van publieke managers om de input ervan te gebruiken (Migchelbrink & Van de Walle, 2019); tussen Public Service Motivation (PSM) en prestaties (Bellé, 2013); tussen PSM en budgettaire strategieën (Moynihan, 2013); tussen publieke leiderschapsstijlen en prestaties (Bellé, 2014); tussen prestatie-informatie over publieke diensten en de perceptie van burgers (James & Moseley, 2014; Olsen, 2017); tussen prestatie-informatie en politieke beslissingen (Desmidt & Meyfroodt, 2020; George et al., 2020; Nielsen & Baekgaard, 2015) tussen administratieve lasten en publieke steun voor overheidsprogramma's (Keiser & Miller, 2020); en tussen administratieve lasten en discriminatie door publieke versus private zorgverstrekkers (Jilke et al., 2018). Elke maand komen er wel nieuwe experimenten en nieuwe thematische invalshoeken bij.

Naast en samen met onderzoekers maken ook beleidsmakers steeds meer gebruik van experimenten en quasi-experimenten om beleidsmaatregelen te evalueren op hun effectiviteit. Door de situatie wanner er geen nieuw beleid (policy off) zou zijn te vergelijken met de bewuste beleidsingreep (policy on) kan worden vastgesteld en beoordeeld of beleid werkt of niet (Van der Knaap et al., 2020). De eerste beleidsexperimenten vonden al plaats in de jaren 1920 en 1930 met *The Lanarkshire School Milk* experiment als één van de bekendste voorbeelden (John, 2017b). Daarna verdwenen ze weer van de radar tot de jaren 1960, op een moment dat de uitrol van omvangrijke, sociale beleidsprogramma's en -hervormingen vroegen om een adequate evaluatie (Campbell, 1969; Greenberg et al., 2003). Vandaaruit breidde het aantal experimenten en de range van beleidsdomeinen steeds verder uit richting arbeidsmarkt (Bloom, 2005), onderwijs (Torgerson, 2001), criminaliteitspreventie (Sherman & Weisburd, 1995), ontwikkelingshulp (Duflo et al., 2007) en belastingen (Hallsworth, 2014).

De vraag die we ons kunnen stellen is: wat maakt experimenten zo aantrekkelijk voor onderzoek naar, voor en door de overheid? Zoals reeds aangehaald schuilt de belangrijkste verklaring in het oplossingspotentieel voor het causaliteitsvraagstuk (Antonakis et al., 2010). Jones stelde al in 1985 (p. 282) dat "experimenten de meest robuuste techniek is die er bestaat om een causale relatie tussen twee variabelen aan te tonen". Andere onderzoekers noemden experimenten zelfs de "gouden standaard" van wetenschappelijk onderzoek (Antonakis, 2017; Eden, 2017; Hauser et al., 2017). Verder biedt de experimentele methode bestuurskundige onderzoekers het potentieel om theoretische hypothesen te toetsen op hun significantie en onderzoekers ten minst een ja/neen antwoord op hun onderzoeksvraag te leveren. Bovendien is het design van een experiment relatief eenvoudig en kan de methode snel en op een bevattelijke manier worden uitgelegd aan leken, practici of studenten (John, 2017c). En niet enkel het design, ook de analyse is relatief overzichtelijk. Om vast te stellen of en wat het effect is op basis van een vergelijking tussen twee groepen, volstaat het om de gemiddelde scores aan een statistische significantietest te onderwerpen. In het algemeen geldt: hoe eenvoudiger het design, hoe eenvoudiger de analyse. Zoals we reeds aangaven, kan de complexiteit van een experiment snel toenemen, zowel qua design als qua analyse (Gerber & Green, 2012). Zelfs bij meer ingewikkelde designs blijft het principe, met name de uitkomsten van willekeurig ingedeelde groepen met elkaar vergelijken, behoorlijk herkenbaar.

De wereldwijde opmars van gedragsteams en nudge-eenheden binnen overheidsinstellingen heeft de experimentele beweging een extra boost gegeven (Afif et al., 2018; John, 2018). Geïnspireerd op de methoden uit de gedragseconomie groeide de RCT uit tot de dominante, empirische methode (Congdon & Shankar, 2018; Mukherjee & Giest, 2020). De voortrekkersrol was weggelegd voor het Britste *Behavioural Insights Team* (BIT), dat het experimenteel testen van het effect van nudges op het individuele gedrag van burgers in het hart van haar werking plaatste (John, 2014). BIT directeur David Halpern meende zelfs dat: "the experimental methods that behaviourists have helped popularise may prove their greatest legacy of all" (2015: p. 339). De visie het Britste BIT werd treffend geïllustreerd in het toonaangevende draaiboek *Test, Learn, Adapt* (Haynes et al., 2012), dat beleidsmakers tot in het detail uitlegt hoe ze een RCT moeten opzetten en evalueren. De achterliggende redenering is dat gedragsinterventies waarvan de effectiviteit op basis van een experimenteel design bij een beperkte steekproef bewezen wordt, vervolgens kunnen opgeschaald worden en breed uitgerold bij de volledige populatie (Al-Ubaydli et al., 2021; Halpern & Mason, 2015).

Ondertussen zouden er al meer dan 400 RCT's op het Britste BIT palmares staan (Sanders et al., 2018), waaronder een aantal bekende veldexperimenten met gedragsinterventies om burgers te helpen hun belastingen tijdig te betalen (Hallsworth et al., 2017), sneller een job te vinden (The Behavioural Insights Team, 2016), energiezuiniger om te springen met elektrische toestellen (The Behavioural Insights Team, 2014) of niet te snel te rijden (BIT UK, 2017). Ook in België zien we de laatste jaren een opmars van beleidsexperimenten op basis van gedragsinzichten, bijvoorbeeld voor het verminderen van de no show bij doktersafspraken (Versavel & Kouwenhoven, 2020) of het stimuleren van gezonde en duurzame voedingskeuzes (Vandenbroele et al., 2018). RCT's zijn overigens allesbehalve nieuw. Zo bewijzen ze al langer hun nut binnen het kader van ex ante en ex post beleidsevaluatie (De Francesco & Pattyn, 2021). Het recentere pleidooi voor RCT's vindt aansluiting bij een rijkere en ruimere tendens rond *evidence-based policymaking* (EBPM), die streeft naar een solide wetenschappelijke onderbouw van beleid (Cairney et al., 2016; Einfeld, 2019; Legrand, 2012). Ook de zogenaamde *what works* momevent kunnen we in datzelfde licht interpreteren (Halpern, 2015; John, 2020).

2.4 HOE EXPERIMENTEREN? EEN STAPPENPLAN

De link tussen gedragsinterventies en RCT's heeft geleid tot een toename van beleidsexperimenten en een toenadering tussen wetenschap en beleidspraktijk (John, 2017b). Beide partijen hebben er immers belang bij om de krachten te bundelen. Beleidsmakers doen een beroep op de expertise van onderzoekers bij het bedenken van interventies en het uitwerken van het experimenteel design. Onderzoekers hebben publieke managers nodig om hun theoretische hypothesen via een overheidsorganisatie te testen in het veld. Onderzoekers en beleidsmakers moeten met elkaar praten, van gedachten wisselen en elkaars denk- en werkwijzen trachten te begrijpen. Het onderzoeksproces en het beleidsproces moeten tot een zekere hoogte op elkaar worden afgesteld. Peter John formuleerde het als volgt: "The partnership is two-way with reserarchers thinking like policy-makers and policy-makers thinking like reserearchers for mutual benefit" (2017c, p. 11).

John onderscheidde 10 cruciale stappen plannen en ontwerpen van veldexperimenten (2017c). De sequentie van het stappenlan is logisch, maar tegelijk zijn de verschillende stappen ook sterk met elkaar verbonden. Een verandering, vergetelheid of vergissing bij de ene stap kan belangrijke gevolgen hebben voor één van de andere stappen. In wat volgt geven we deze 10 stappen weer, met een korte samenvatting per stap. We vermelden hier graag uitdrukkelijk dat we ons voor dit luik in grote mate laten inspireren op het werk van Peter John (2017c, pp. 17–48). Het stappenplan werd opgevat en opgemaakt vanuit het perspectief van de onderzoeker, maar heeft ook oog voor het perspectief van de beleidsmaker.

Stap 1: Fit for trial? Nagaan of een experiment wenselijk en haalbaar is

Niet alle onderzoeksvragen of beleidsproblemen kunnen beantwoord of opgelost worden door middel van een experiment. De eerste vraag die we ons moeten stellen is of een experiment wel de aangewezen methode is. En zelfs mocht dat het geval zijn, dan nog moeten we bekijken of het experiment uitvoerbaar en wenselijk is. Zijn er wel voldoende onderzoekseenheden? Is het ethisch te verantwoorden? Kan er voldoende gerandomiseerd worden? Staat het datamanagement op punt? (Glennerster & Takavarasha, 2013).

Stap 2: Het rekruteren van een partner

Wanneer het onderzoek zich op de interactie tussen overheid en burgers richt, zal een veldexperiment doorgaans gebeuren op basis van een samenwerking tussen academische onderzoekers en externe partners, in casu overheidsorganisaties. Dit partnerschap kan onderdeel zijn van een officieel bestek, bijvoorbeeld in het kader van een evaluatie-opdracht. In het geval van nudge-experimenten wordt er ook soms gewerkt op basis van informele contacten via studiedagen, opleidingen en netwerken (John, 2017c). Het engagement van de partner is cruciaal in dit verhaal en dient op papier geofficialiseerd te worden. Dit engagement wordt idealiter uitgedrukt doordat de organisatie budget, middelen en mensen vrijmaken voor het experiment. Daarnaast is het belangrijk om van bij het begin duidelijkheid te scheppen over de verwachtingen, de doelstelling, de noden en de kosten.

Stap 3: Het bepalen van de onderzoeksite en steekproef

In de realiteit zal de locatie en populatie van het onderzoek samenhangen met de (keuze van) partner. Ook de selectie van de onderzoekseenheden zal bij een veldexperiment in grote mate bepaald worden door de actieradius en processen van de partner. Binnen het onderzoeksveld van *Behavioural Public Policy & Administration* wordt meestal gekozen voor individuele personen zoals burgers, ambtenaren, beleidsmakers of politici. Binnen bestuurskundig onderzoek worden onderzoekseenheden ook soms op een hoger niveau ingedeeld, zoals teams, gemeenten of organisaties (van Thiel, 2015).

Stap 4: Het uitwerken van de interventies

De interventies vloeien idealiter logisch voort uit het theoretische model van de onderzoeker of uit de beleidsmaatregel die de beleidsmaker wil evalueren. Toch wordt men in de realiteit vaak geconfronteerd met praktische beperkingen en moet men op zoek naar de juiste balans tussen interventies die theoretisch geldig blijven, maar ook op een realistische en haalbare manier toegediend kunnen worden. Bovendien is het belangrijk dat elk interventie slechts één mechanisme of component uit het theoretisch model test, anders is het onmogelijk om te evalueren welk afzonderlijk element nu net een bepaalt effect genereert. Daarbij dient beslist te worden hoeveel interventies en interventiegroepen men wil testen. Indien de steekproefgrootte en dus het aantal mogelijke interventiegroepen beperkt is, wordt aangeraden om te focussen op één of twee cruciale interventies.

Stap 5: Het vastleggen van de controlegroep

De controlegroep is essentieel omdat deze informatie oplevert over de zogenaamde policy off situatie, wanneer er geen (nieuwe) beleidsmaatregelen of gedragsinterventie zou zijn. Bij experimenten spreken we in dat geval van de *counterfactual*. Bij nudge-experimenten met brieven wordt voor de controlebrief vaak gekeken naar de originele brief die al verstuurd wordt (zie Hoofdstuk 4). Idealiter bezitten de interventie van de controlegroep en de andere interventies een gemeenschappelijke basis, zodat het mogelijk wordt om het effect van het toegevoegde element te meten. Ook het gebruik van placebo's kan hierbij handig zijn, maar is in de sociale wetenschappen minder gebruikelijk en minder evident dan in de medische wetenschappen.

Stap 6: Het berekenen van de steekproefgrootte

Een cruciale vraag de in elk experiment aan bod komt is: hoe groot moet de steekproef zijn? Het doel is in ieder geval te beschikken over een steekproef die groot genoeg is om met enige betrouwbaarheid het effect te kunnen vaststellen van een interventie ten opzichte van een controlegroep of tussen interventies onderling. In principe geldt de volgende richtlijn: hoe groter de steekproef, hoe kleiner het te detecteren effect. Bij een grote(re) steekproef kunnen echter ook extra kosten komen kijken, waardoor het toch interessant kan zijn om de steekproefgrootte enigszins te kalibreren. Om dit te berekenen heeft de onderzoeker op voorhand informatie nodig over de verwachtte effectgrootte en spreiding. Daarbij is het ook belangrijk te kijken naar de kracht van een statistische toets, die bepaalt hoe groot het risico is dat een onderzoeker een vals positief effect vindt (de standaardmaat voor power bij een experiment in de sociale wetenschappen =0.80). Een algemeen advies bij het bepalen van de steekproefgrootte is om nogal pessimistisch en conservatief ingesteld te werk te gaan. Beter een te grote steekproef dan een te kleine steekproef.

Stap 7: Het scherpstellen van het randomisatieproces

Zoals hierboven geschetst is de randomisatie één van de drie cruciale criteria om te spreken van een zuiver experiment. Het ontwerp van een randomisatieproces is op zichzelf relatief eenvoudig en kan gebeuren via statistische programma's, specifieke software of een Excel document. Bij de implementatie komt echter meer kijken. Vaak zal het immers de overheidsorganisatie zijn die de randomisatieprocedure tijdens de

volledige duur van het experiment moet toepassen en opvolgen. De medewerkers moeten het principe en het belang van randomisatie begrijpen en het protocol rigoureus volgen.

Stap 8: Het plannen van de dataverzameling

Een experiment staat of valt met de kwaliteit van de data. Een experiment heeft eerste en vooral nood aan een duidelijk datamanagementprotocol, dat rekening houdt met de privacyrichtlijnen en duidelijkheid schept over de periode waarin en de manier waarop de data verzameld, uitgewisseld en opgeslagen wordt. De dataset wordt standaard geanonimiseerd. Naast informatie betreffende de afhankelijke variabelen moeten ook de zogenaamde covariaten zoals bijvoorbeeld leeftijd, geslacht of opleiding gekozen worden. Deze covariaten gebruiken we ook om na te gaan of de groepen in balans zijn.

Stap 9: Het inschatten van de ethische principes

Onderzoekers moeten rekening houden met een reeks ethische principes bij het uitvoeren van veldexperimenten. Ze mogen deelnemers aan experimenten met respect behandelen en mogen hen op geen enkele manier benadelen. Wanneer het gaat om kwetsbare personen wegen deze criteria nog zwaarder door. Geïnformeerde toestemming is het streefdoel, maar is bij veldexperimenten niet altijd mogelijk. In dat geval kan het voorzien van een debriefing waarbij de deelnemers achteraf alsnog ingelicht worden, soelaas bieden (McClendon, 2012). Misleiding moet vermeden worden (Colson et al., 2016). De informatie die gecommuniceerd wordt, dient waarheidsgetrouw te zijn. Voor onderzoekers geldt dat voor deze experimenten goedkeuring gevraagd moet worden aan een ethische commissie. De twee experimenten in dit rapport werden allebei voorgelegd aan goedgekeurd door de Sociaal-Maatschappelijke Ethische Commissie van de KU Leuven. Beleidspractici zijn niet verplicht om via een ethische commissie te passeren, maar moeten zich minstens verzekeren van politieke en ambtelijke steun.

Stap 10: Het voorbereiden van de onderzoeksplannen

In laatste instantie wordt alles best neergepend in een overkoepelend onderzoeksplan dat voor alle partijen beschikbaar en helder is. Daarnaast gaan sociale wetenschappers vandaag vaker werken via (pre)registratie van de onderzoeksplannen. Vooraleer het experiment start worden de variabelen, interventies, procedures en verwachte uitkomsten op een publiek toegankelijk platform opgeslagen zodat er nadien minder kans is

op *p-value fishing* naar en *cherry picking* van statistisch significante resultaten (James, Jilke, et al., 2017b). Tot slot bestaan er ook richtlijnen inzake de rapportering van experimenten, zie bijvoorbeeld ook het CONSORT diagram dat bestaat uit 25 elementen om rekening mee te houden (John, 2017c)

2.5 WAAR LIGGEN DE EXPERIMENTELE GRENZEN?

Experimenten op basis van RCT's vormen een krachtige methode om overheidsbeleid te toetsen en te versterken. In de realiteit botsen dit soort beleidsexperimenten echter dikwijls op ethische, praktische en methodologisch limieten. Bij de negende stap van het stappenplan stonden we al stil bij de ethische principes die gerespecteerd moeten worden. Denk aan het respectvol behandelen en eerlijk informeren van deelnemers, het garanderen van de privacy, het streven naar geïnformeerde toestemming en de goedkeuring van een ethische commissie. Daarenboven moeten onderzoekers er zich van bewust zijn dat hun (veld)interventies ingrijpen in het leven van mensen. Door een interventie aan de ene groep wel en aan de andere groep niet toe te dienen kunnen bepaald voordelen ontzegd of nadelen gecreëerd worden, wat een gevoel van discriminatie en onrechtvaardigheid kan veroorzaken. Zoiets kan deels verholpen worden door te werken met een soort wachtlijst of een *stepped wedged* design, waarbij alle deelnemers na verloop van tijd stap voor stap toegang tot de interventie krijgen (Gerber & Green, 2012, pp. 276–281). Zeker indien er gewerkt wordt met kwetsbare groepen zoals bijvoorbeeld mensen in armoede, werklozen of personen met een handicap, dienen bovenstaande ethische principes extra aandacht te krijgen (James, John, et al., 2017).

Naast de ethische principes, hebben experimenten ook een hele reeks praktische en methodologische pijn- en aandachtspunten. In Box 1 bieden we een overzicht van 9 veelvoorkomende implementatieproblemen bij het uitvoeren van een experiment. Er kan bijvoorbeeld iets misgaan bij de randomisatie, de rekrutering van deelnemers, het toedienen van interventies, het verzamelen van de data en het analyseren van de data. Deze lijst is allesbehalve exhaustief. In het algemeen loopt een experiment immers zelden zoals het op voorhand gepland was en op papier stond. Onverwachte obstakels liggen in de lijn der verwachtingen. Onderweg zal er permanente monitoring en bijsturing nodig zijn. Onderzoekers met voeling voor en ervaring met adaptief projectmanagement hebben hier een stapje voor. Ook het métier van de publieke (project)managers die bij het experiment betrokken zijn, komt hier goed van pas.

Problemen bij het implementeren van beleidsexperimenten

Probleem 1: moeite met het rekruteren van deelnemers: de responsgraad is lager dan verwacht, deelnemers stappen uit het experiment, wat problemen kan opleveren voor de statistische power en bias kan veroorzaken wegens verloop

Probleem 2: mislukte of beschadigde randomisatie: dit kan het gevolg zijn van gebrekkige samenwerking tussen de partners, een misverstand of een fout

Probleem 3: interventie die niet of niet zoals bedoeld toegediend is: dit heeft te maken met de praktische implementatie waar technische en menselijke fouten kunnen gebeuren.

Probleem 4: de controlegroep ontvangt een interventie: de interventies worden niet aan de gepland groep bezorgd (crossover).

Probleem 5: contact tussen interventie- en controlegroep: beide groepen maken contact met elkaar en beïnvloeden elkaar (contaminatie).

Probleem 6: onverwachte schokken: gebeurtenissen kunnen van buitenaf komen zoals overstromingen of pandemieën, maar kunnen ook intern voorvallen zoals fusies, hervormingen of de introductie van nieuwe programma's.

Probleem 7: slechte kwaliteit van data: een veldexperiment is in grote mate afhankelijk van de kwaliteit van de monitoringsystemen en dataprocessen van de betrokken overheidsorganisatie. Vaak is deze organisatie niet vertrouwd met de dataformats die nodig zijn, waardoor data kan verloren gaan. Het matchen van de interventiedata en de effectdata behelst vaak het samenvoegen van twee bestanden en is een delicate oefening.

Probleem 8: verlies van samenwerking van of tussen partners: naarmate het experiment vordert krijgen partners zicht op de kosten en inspanningen die nodig zijn en kunnen ze hun vertrouwen en interesse in het project verliezen.

Probleem 9: onverwachte logistieke uitdagingen en kosten: het betreft een restcategorie van alle zaken die kunnen fout lopen bij de implementatie, zeker wanneer experimenten boven het budget en timing dreigen te gaan.

Box 1. Problemen bij het implementeren van beleidsexperimenten (John, 2017)

En zelfs al omzeilen of ontmantelen onderzoekers deze potentiële problemen en brengen ze een experiment tot een goed einde, dan nog moeten we ons afvragen wat deze resultaten betekenen voor theorie en praktijk. De specifieke context van een interventie en de algehele complexiteit van de samenleving moeten ons hoeden voor een overschatting van onze bevindingen en een al te sterk geloof in eenduidige, universele en definitieve antwoorden. Anders gesteld, RCT's vertellen meer iets over "wat werkte" in een specifieke setting, dan over "wat altijd en overal werkt" (Biesta, 2007). Replicatiestudies, het herhalen van experimenten in eenzelfde of andere context, kan deze tekortkoming deels verhelpen door de externe validiteit te verhogen (Schmidt, 2009; Walker et al., 2017).

Naast overschatting willen we hier ook nog een keer wijzen op de risico's van methodologische vernauwing of methodolotry (Hassan & Wright, 2020; Moynihan, 2018). Methodologische modeverschijnselen zijn tot op zekere hoogte onvermijdbaar en hoeven niet noodzakelijk problematisch te zijn. Wanneer onderzoekers en evaluatoren in hun zoektocht naar interne validiteit en de bijhorende empirische geloofwaardigheid echter steeds opnieuw de voorkeur geven aan onderzoeksonderwerpen die zich qua methode (enkel) lenen tot een experimenteel design, dreigen veel bestuurskundige onderzoeksvragen gestuurd te worden op basis van methodologische preferenties en trends (Van de Walle, 2017). Zo zien we bijvoorbeeld nu al een verschuiving van de werking van instellingen naar het gedrag van individuen (Bertelli & Riccucci, 2020; Grimmelikhuijsen et al., 2017). Ook complexere thema's en processen zouden op die manier in de achtergrond verzeilen. Daarom blijft het belangrijk om op methodologisch vlak te variëren en verschillende technieken te combineren, teneinde theoretische kennis te cumuleren (Kinder, 2011; Van de Walle, 2017). Op die manier verhogen we onze kansen om niet alleen te achterhalen "wat werkte", maar ook "waarom iets werkte" (Deaton & Cartwright, 2018).

3 SURVEY-EXPERIMENT SNELHEIDSBOETES

Overdreven of onaangepaste snelheid is een fundamentele oorzaak van verkeersonveiligheid. Snelheid beïnvloedt zowel de kans op ongevallen als de ernst ervan (Stradling, 2000; Taylor et al., 2000). Te hoge snelheid ondermijnt reactietijd, remafstand en controle over het voertuig (Aarts & Van Schagen, 2006). Naar schatting zijn 10% tot 15% van alle ongevallen en 30% van de ongevallen met dodelijk letsel het rechtstreekse gevolg van overdreven of onaangepaste rijsnelheid (Van den Berghe & Pelssers, 2020). Ondanks de overduidelijke risico's verbonden aan overdreven snelheid overschrijden bestuurders de snelheidslimieten nog steeds veelvuldig, zoals blijkt uit het hoge aantal verkeersovertredingen in België en uit de resultaten van nationale en internationale attitudemetingen aantonen (Holocher & Holte, 2019; Yannis et al., 2016). Als de overheid erin slaagt om rijgedrag in het algemeen positief te beïnvloeden, en om meer specifiek ook in te grijpen op het beslissingsproces rond rijsnelheid, kan dit de veiligheid in het verkeer voor iedereen verhogen. In de problematiek rond rijsnelheid is de snelheidsboete een alom gekend en beproefd beleidsinstrument (Chen et al., 2020; du Plessis et al., 2020; Hössinger & Berger, 2012). Het voorbije decennium werden in België ieder jaar meer dan drie miljoen snelheidsovertredingen vastgesteld (Van den Berghe & Pelssers, 2020). Sinds 2019 komen we zelfs uit boven de vier miljoen snelheidsboetes (Federale Politie, 2021)[2].

Op het eerste gezicht lijkt een boete niet veel meer dan een administratieve en financiële interactie die de burger erop attent moet maken dat hij of zij de (verkeers)wet heeft overtreden en hiervoor bestraft wordt (Hössinger & Berger, 2012). Boetes vormen echter ook een zogenaamd *touch point* of een contactmoment tussen overheid en burger. Vanuit dit perspectief bieden boetebrieven die overheid een ingang om burgers te informeren, sensibiliseren en hun gedrag te beïnvloeden. In deze studie testen we het potentieel van zulke persuasieve boodschappen, die in de vorm van een bijlage toegevoegd worden aan de boetebrief.

Sinds 2019 worden bij een aantal verkeersboetes extra fiches of documenten toegevoegd, zoals informatie over de wetgeving of de risico's verbonden aan het begaan van een verkeersinbreuk, denk bijvoorbeeld aan te snel rijden, gsm-gebruik achter het stuur of alcohol. Deze mogelijkheid is er gekomen doordat de print- en verzendkosten bij Bpost voor een extra pagina bij de verkeersboete dezelfde zijn. De FOD Justitie

[2] zie http://www.verkeersstatistieken.federalepolitie.be/verkeersstatistieken/interactief/

(bevoegd voor het verdelen en innen van de verkeersboetes) gaf daarom aan de FOD Mobiliteit de kans om hier gebruik van te maken. Aangezien snelheidsovertredingen buiten snelwegen een gewestelijke bevoegdheid zijn, laat de FOD Mobiliteit voor die overtredingen de mogelijkheid aan Vlaanderen om een extra fiche bij te voegen. De inhoud en het ontwerp van deze bijlage is in handen van de Vlaamse Stichting Verkeerskunde (VSV), die in opdracht van het Departement Mobiliteit en Openbare Werken (DMOW) de sensibilisering rond verkeersveiligheid vormgeeft.

Voor dit experiment testen we samen met de VSV de effectiviteit van een rationele en een emotionele boodschap op de intentie van autobestuurders om zich aan de snelheidslimieten te houden. Op basis van een gerandomiseerd survey-experiment (n=600) evalueren we twee soorten boodschappen die beiden op het terrein als bijlage bij de boetebrieven toegevoegd werden of worden. Dit survey-experiment kadert tevens binnen de bredere evaluatie van de campagne "Een beetje te snel is ook te snel", die door VSV in het najaar van 2020 werd georganiseerd. In Box 2 geven we als samenvatting de belangrijkste elementen van deze case mee op basis van het stappenplan voor het opstellen van een gedragsinterventie (zie Hoofdstuk 1).

Stappenplan voor het opzetten van gedragsinterventies: case rijsnelheid

Stap 1: Definitiefase

Wat is het probleemgedrag? Autobestuurders houden zich niet (altijd) aan de wettelijke snelheidslimieten. Wat is het gewenste gedrag? Autobestuurders respecteren de snelheidslimieten. Wie is de doelgroep? In deze studie werken we met een steekproef voor de gehele Vlaamse populatie van autobestuurders tussen 18 en 54 jaar. We focussen niet specifiek op één doelgroep op basis van leeftijd en geslacht (bv. jonge mannen), maar houden in onze analyses wel rekening met mogelijke verschillen. Verder is belangrijk om te vermelden dat we niet enkel autobestuurders met een recente snelheidsovertreding bestuderen, maar alle bestuurders.

Stap 2: Analysefase

Wat zijn de determinanten om te snel te rijden? Te snel rijden kan zowel bewust als onbewust gebeuren en wordt beïnvloed door externe factoren (weginfrastructuur, autotechnologie) en interne factoren (attitude, sociale normen of risicoperceptie). In deze studie nemen we de *Protection Motivation Theory* als gedragsmodel, met dreiging en coping als de twee bepalende cognitieve processen.

Stap 3: De interventiefase

We testen het effect van rationele versus emotionele boodschappen op de intentie om te snel te rijden. Deze boodschappen worden als bijlage toegevoegd aan een fictieve, administratieve boetebrief. De rationele interventie werd in het verleden ontworpen door VSV en gebruikt in het veld. De emotionele interventie is geïnspireerd op een veldexperiment van het *Behavioural Insights Team* (2017) in het Verenigd Koninkrijk en wordt vandaag gebruikt als bijlage bij een snelheidsovertreding in Vlaanderen.

Stap 4: De designfase

We maken gebruik van een survey-experiment (n=600) en verdelen de respondenten op basis van individuele randomisatie in drie groepen: een controlegroep, die enkel een administratieve brief te zien krijgt; een eerste interventiegroep, die de administratieve brief en de rationele bijlage te zien krijgt en de tweede interventiegroep, die. de administratieve brief en de emotionele bijlage te zien krijgt We meten de intentie zich aan de rijsnelheid te houden zowel voor als na dat de deelnemers de brieven zien.

Stap 5: De evaluatiefase

Op basis van gepaarde t-toetsen vergelijken we binnen elk van de drie groepen de gemiddelde intentie om te snel te rijden voor en na het zien van de brieven en gaan na of er significante effecten zijn. Daarnaast stellen we ook een aantal appreciatievragen over de duidelijkheid, geloofwaardigheid en herkenbaarheid van de brieven. Deze resultaten zetten we om in concrete beleidsaanbevelingen.

Box 2. Samenvatting case rijsnelheid volgens stappenplan

3.1 DE GEDRAGSFACTOR IN HET VERKEER

Menselijk gedrag en verkeersveiligheid zijn onlosmakelijk met elkaar verbonden. Bewust en onbewust snelheidsgedrag stelt zich via een complex proces dat door verscheidene, vaak samenhangende factoren wordt beïnvloed. Enerzijds zijn er externe factoren zoals de weginfrastructuur (bv. leesbare wegen) en de steeds sneller evoluerende autotechnologie (bv. intelligente snelheidsbegrenzing – ISA). Anderzijds blijven interne, menselijke mechanismen zoals attitude, sociale normen of risicoperceptie doorslaggevend in het stellen van snelheidsgedrag. De interdependentie van deze externe en interne factoren maakt het bijzonder uitdagend om rijgedrag te beïnvloeden. Niettemin weten we dat op theorie en empirie gebaseerde antisnelheidscampagnes en interventies effectief kunnen zijn en dat zulke campagnes zowel intentie als gedrag inzake rijsnelheid kunnen beïnvloeden (Wundersitz et al., 2010). Het is essentieel om te weten welke instrumenten en interventies kunnen werken en het rijgedrag beïnvloeden. Onderzoek heeft bijvoorbeeld aangetoond dat snelheidscamera's en een slimme (her)inrichting van de rij-infrastructuur, onder meer via vluchtheuvels en rotonden, het meeste potentieel hebben om ongevallen te verminderen (Bunn et al., 2003; Pilkington & Kinra, 2005). Deze beleidsinstrumenten kunnen aangevuld worden met instrumenten die zich richten op de motivatie en attitudes van autobestuurders teneinde hun rijgedrag aan te passen.

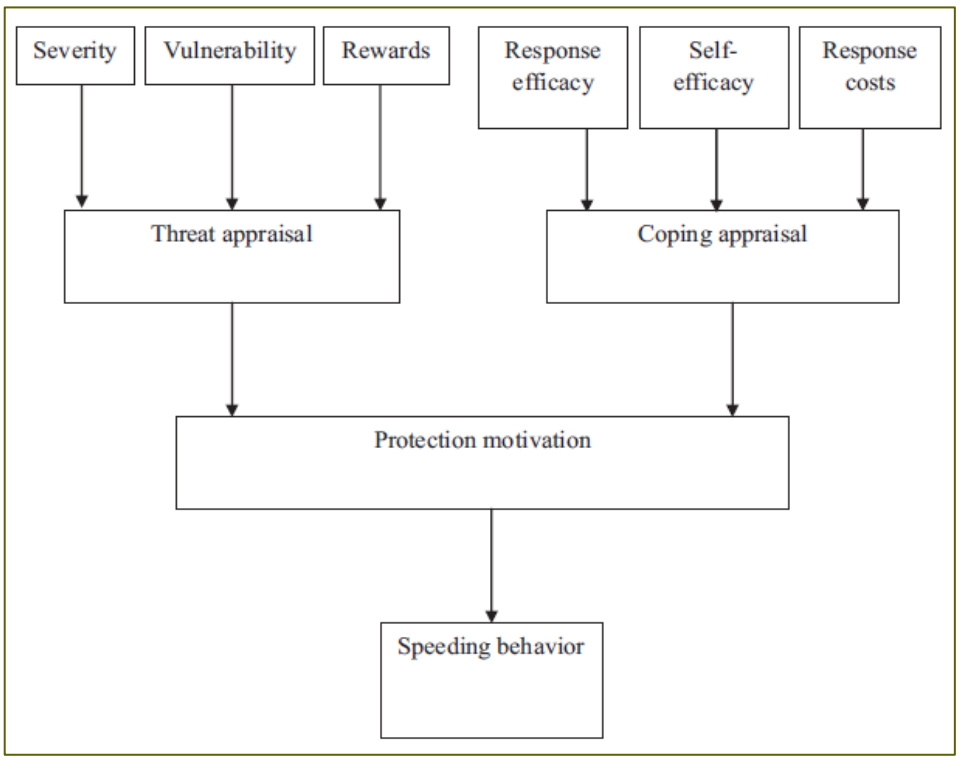
3.1.1 Protection Motivation Theory

Gezien de centrale plaats van motivatie, attitudes en gedrag in verkeersveiligheidsbeleid- en onderzoek hoeft het niet te verbazen gedragsinzichten er al geruime tijd prominent onder de aandacht staan. De veelvuldig gebruikte en gerepliceerde *Driver Behavior Questionnaire* (DBQ) is een uitstekende illustratie van gedragswetenschappelijk verkeersveiligheidsonderzoek (Reason et al., 1990). Deze vragenlijst wordt gebruikt om (de intentie tot) 'afwijkend' rijgedrag te testen en maakt een onderscheid tussen bewuste overtredingen en onbewuste vergissingen (Özkan et al., 2006; Parker et al., 1996). Op basis van DBQ onderzoek blijkt bijvoorbeeld dat mannen en jongere chauffeurs vaker overtredingen begaan dan vrouwen, en dat oudere of professionele chauffeurs en bestuurders van bedrijfswagens de regels vaker overtreden. Verkeersveiligheidsstudies doen bovendien steeds vaker een beroep op algemene theoretische gedragsmodellen zoals de *Theory of Planned Behavior* (Ajzen, 1991; Conner et al., 2007; Paris & Van den Broucke, 2008; Warner & Åberg, 2006); het *Elaboration Likelihood Model* (Lewis et al., 2008; Petty & Cacioppo,

1986) het *Extended Parallel Process Model* (Lewis et al., 2009; Witte, 1992) of de *Protection Motivation Theory* (Glendon & Walker, 2013; Maddux & Rogers, 1983).

Deze studie baseert zich op de *Protection Motivation Theory* (PMT), omdat dit model zowel toegepast kan worden op het design én de evaluatie van rationele en emotionele boodschappen rond verkeersveiligheid (Glendon & Walker, 2013). Oorspronkelijk werd de PMT ontwikkeld om de effecten van bedreigende of angstaanjagende boodschappen op preventief gezondheidsgedrag te verklaren (Floyd et al., 2000). Het model bestaat uit twee processen: dreiging (*threat appraisal*) enerzijds en coping (*coping appraisal*) anderzijds. De component dreiging schat de gevolgen en risico's van het ongewenste gedrag in (c.q. te snel rijden) en is opgebouwd uit drie factoren: een gevoel van beloning (*rewards*), de inschatting van de ernst van de dreiging (*severity*) en de perceptie van de eigen kwetsbaarheid ten aanzien van de dreiging (*vulnerability*). Een toename op het vlak van beloning zal de kans op het ongewenste gedrag doen toenemen, terwijl ernst en kwetsbaarheid de kans op te snel rijden doen afnemen. De coping component heeft betrekking op de manier waarop men omgaat met de dreiging en de strategieën om de dreiging af te wenden (c.q. de snelheidslimieten te respecteren). De coping component bestaat eveneens uit drie factoren: de potentiële kosten (niet enkel financieel, maar ook materieel, sociaal en psychologisch) die verbonden zijn met het gewenste gedrag (*costs*), de mate waarin men eenvoudig het gewenste gedrag kan stellen (*self-efficacy*) en de inschatting van de effectiviteit van aangereikte antwoorden om het gewenst gedrag te vertonen en zo de dreiging af te wenden *(response efficacy)*.

De twee processen van dreiging en coping worden weergegeven in Figuur 6 en vormen samen de *protection motivation*, zijnde de motivatie om zich te beschermen tegen dreiging. *Protection Motivation* is de uitkomst van beide processen en wordt vaak weergegeven en gemeten op basis van intentie (Floyd et al., 2000; Glendon & Walker, 2013).



Figuur 6. Protection Motivation en Speeding Behaviour (Glendon & Walker, 2013, p. 69)

3.1.2 Nudging en rijsnelheid

Naast de uitgebreide toepassing van gedragstheorieën en gedragsmodellen merken we dat recent ook de term nudging binnen verkeersveiligheidsonderzoek aan belang wint. Het idee achter nudging is echter al veel langer ingeburgerd, samen met de bredere toepassing van inzichten uit de sociale psychologie en gedragseconomie. Graf (2019) spreekt vanuit die optiek over 'nudging before the nudge' en illustreert dit op treffende wijze door de historische interventies ter promotie van het gordelgebruik te beschrijven. Ook de gedragsinterventies op de Lake Shore Drive (LSD) in Chicago, beschreven in het Nudge boek van Thaler en Sunstein (2008), gaan meer dan vijftien jaar terug in de tijd. Bij de Lake Shore Drive werden de dwarslijnen in de aanloop naar een gevaarlijke bocht steeds dichter bij elkaar geplaatst zodat de bestuurders uiteindelijk hun snelheid minderden en de kans op een ongeval afnam. De voorbije jaren werden er een hele reeks nudge-interventies ontwikkeld op het vlak van verkeersveiligheid en snelheid.

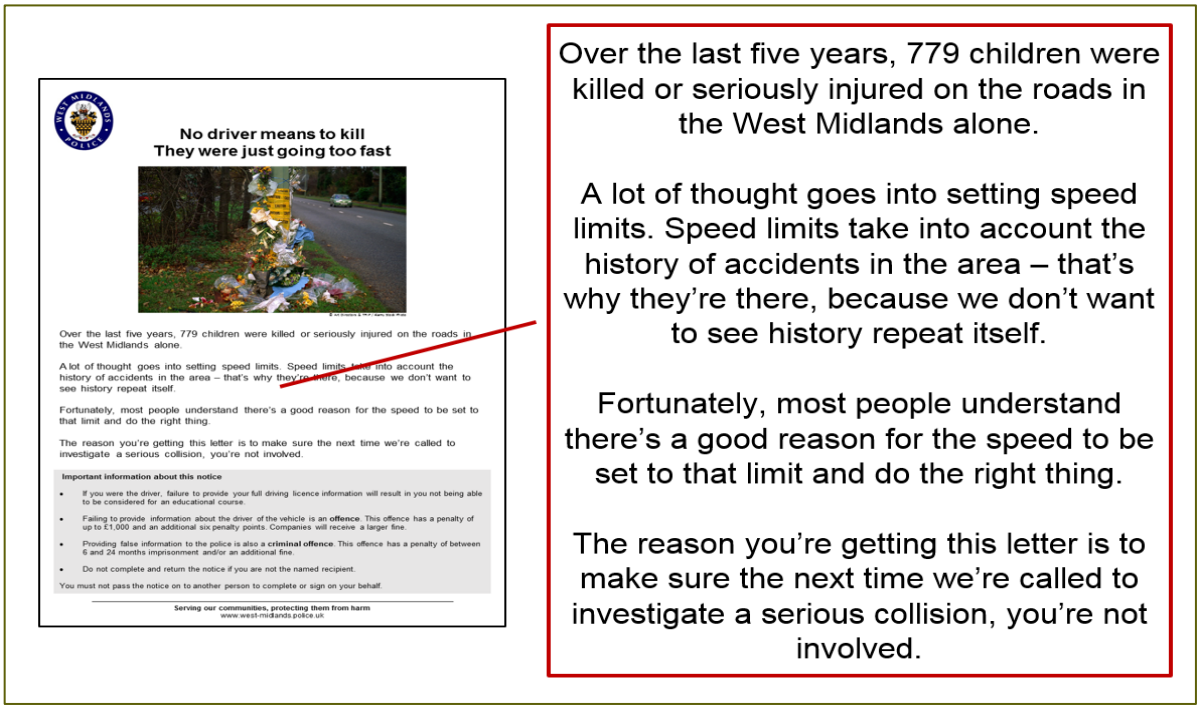
- In Nederland experimenteerde het nationaal wetenschappelijk instituut voor verkeersveiligheidsonderzoek (SWOV) met Dick Bruna-borden langs 30 km/uur-wegen. Bestuurders die deze attentieborden zagen, werden op basis van priming beïnvloed om hun vaart te minderen (Goldenbeld et al., 2017).

- In Zweden werden snelheidscamera's gelinkt aan een loterijsysteem waarbij bestuurders die zich aan de rijsnelheid hielden kans maakten op prijzengeld. Het prijzengeld zelf werd op zijn beurt ingezameld via de boetes van bestuurders die zich niet aan de rijsnelheid hielden (Zichermann & Cunningham, 2011).

- Italiaanse onderzoekers experimenteerden met het *left-digit effect* en vonden dat mensen geneigd waren om trager te rijden als het verkeersbord een snelheidslimiet van 49 km/u aangaf in plaats van 50 km/u (Rubaltelli et al., 2021).

- In België zien we het gebruik van gedragsinzichten en nudging technieken eveneens toenemen. Denk maar aan de "Beloofd" campagnes van de VSV die inzet op sociale normen en precommitment strategieën om mensen ertoe te verbinden niet te snel ter rijden, gordels te dragen of hun gsm niet te gebruiken achter het stuur. De proefprojecten met 3D voetgangersoversteekplaatsen en gevleugelde zebrapaden op verschillende locaties in Vlaanderen onderschrijven die trend (De Ceunynck, Pelssers, & Daniels, 2020; De Ceunynck, Pelssers, Daniels, et al., 2020)

Voor deze studie naar het potentieel van boetebrieven op het rijgedrag van bestuurders hebben we ons laten inspireren door een experiment uit het Verenigd Koninkrijk. Daar ontwierp het *Behavioural Insights Team*, ook wel de *Nudge Unit* genoemd, i.s.m. de *West Midlands Police* een succesvolle beleidsinterventie ten aanzien van snelheidsovertreders (BIT UK, 2017). Het betrof een grootschalig cluster RCT veldexperiment rond boetes voor snelheidsovertreders. Aan de standaardbrief die snelheidsovertreders ontvingen, de *zogenaamde Notice of Intended Prosecution*, werd een confronterende, emotionele bijlage toegevoegd met een afschrikwekkende afbeelding en de boodschap dat de voorbije vijf jaar 779 kinderen omkwamen of zwaargewond geraakten bij auto-ongevallen in de streek. De brief wordt weergegeven in Figuur 7.



Figuur 7. Notice of Intended Prosecution en emotionele bijlage (The Behavioural Insights Team, 2017)

In dit Engelse experiment had de bijlage een dubbel doel. Enerzijds was de bijlage erop gericht snelheidsovertreders te overhalen om akkoord te gaan met de voorgestelde sanctie en op tijd te betalen. Anderzijds werd de bijlage toegevoegd om bestuurders te overtuigen van het nut en de legitimiteit van snelheidslimieten zodat zij in de toekomst niet opnieuw te snel zouden rijden. Het aantal betalingen bij de interventiegroep steeg met 13,7% en er werd 20% sneller betaald. Nog belangrijker was dat over een periode van zes maanden na de overtreding het aantal recidivisten significant daalde van 3,1% in de controlegroep naar 2,45% in de interventiegroep (BIT UK, 2017). Het design en de resultaten van dit experiment in het VK vormen het vertrekpunt voor het experiment in Vlaanderen.

3.2 RATIONELE VERSUS EMOTIONELE BOODSCHAPPEN

Meta-analyses naar verkeersveiligheidscampagnes tonen aan dat persuasieve boodschappen die voortbouwen op wetenschappelijke inzichten niet alleen effectiever zijn dan a-theoretische, intuïtieve campagnes, maar ook beter scoren op het vlak van kostenefficiëntie en evaluatiegemak (Elliott et al., 2003;

Lewis et al., 2009). Toch moeten we vaststellen dat de meeste campagnes op dit terrein *theory* noch *evidence based* zijn (Wundersitz et al., 2010). Er is dus nood aan meer wetenschappelijke inzichten over welke soorten boodschappen effectief zijn, waarom en voor welke doelgroep. Vandaag weten we al dat boodschappen duidelijk, consistent, eenduidig en begrijpelijk moeten zijn voor het doelpubliek. Hoe vaker de boodschap herhaald wordt en hoe meer verschillende kanalen worden ingeschakeld, hoe beter. Boodschappen kunnen in grote lijnen een rationale of een emotionele stijl hebben (Delhomme et al., 2009; Warner & Forward, 2016). Rationele boodschappen benadrukken feitelijke informatie, volgen een deductieve logica en spelen in op reflectieve, cognitieve processen. Ze wijzen op de voordelen van het gewenst gedrag en de nadelen van het probleemgedrag (Wundersitz et al., 2010).

3.2.1 De rationele interventie

De boodschap die door de VSV ontworpen werd en die door de FOD Mobiliteit van oktober 2019 tot december 2020 als bijlage bij een snelheidsboete werd verzonden, bevat een rationele dimensie. De bijlage bestaat uit twee delen. In het eerste deel (zie Figuur 8) worden de voordelen van het respecteren van de snelheidslimieten te houden opgesomd en gevisualiseerd.



Figuur 8. Eerste deel van de rationele bijlage (VSV s.d.)

Door minder snel te rijden, zo geeft de boodschap mee, merk je als chauffeur meer op, vergroot je je gezichtsveld, vergroot je de kans om op tijd te stoppen, en zullen de gevolgen van een aanrijding minder ernstig zijn.

Nadat de voordelen van zich aan de rijsnelheid te houden uiteen zijn gezet, volgt een reeks tips om mensen te helpen dit voordelig gedrag te stellen (zie Figuur 9). De rationele boodschap geeft mee dat je aan de snelheidslimiet houden gemakkelijker wordt door een rit op voorhand uit te stippelen, op tijd te vertrekken, de snelheidsbeperkingen via gps op te volgen, zich niet te laten opjagen en steeds rekening te houden met de omstandigheden.



Figuur 9. Tweede deel van de rationele bijlage (VSV, s.d.)

3.2.2 De emotionele interventie

Emotionele boodschappen volgen een andere werkwijze en roepen bepaalde gevoelens op. Emotionele reacties kunnen zowel positief zijn (liefde, geluk, blijdschap) als negatief (angst, woede, verdriet). Angstaanjagende boodschappen, ook wel *fear based messages* of *fear appeals* genoemd, worden vaak gebruikt in verkeersveiligheidscampagnes (Lewis et al., 2007; Witte & Allen, 2000). Deze boodschappen slagen er doorgaans goed in om de aandacht van mensen te trekken en kunnen, mits aan enkele voorwaarden is voldaan, ook de nodige motivatie leveren om gedrag te veranderen. In het kader van dit onderzoek gaat het bijvoorbeeld om boodschappen die snelheidsovertredingen in verband brengen met verkeersongevallen

en die beelden tonen van materiële en fysieke schade. De bijlage die het BIT UK ontwierp voor het veldexperiment in het Verenigd Koninkrijk is een typevoorbeeld van een emotionele en zelfs schokkende boodschap. De emotionele boodschap bevat een afbeelding van een herdenkingsteken voor een verkeerslachtoffer en een verwijzing naar de kwetsbaarheid van kinderen. De bijlage die we in het kader van dit onderzoek en de algemene snelheidscampagne samen met VSV ontwierpen is subtieler, maar bevat ook een aantal emotionele elementen (zie Figuur 10).



Figuur 10. Emotionele bijlage in deze studie

De emotionele bijlage in deze studie haalt aan dat te hoge snelheid de oorzaak is van 1 op de 3 dodelijke ongevallen. Verder springt de afbeelding van de fiets in het oog die op het wegdek ligt en die de suggestie van een zwaar ongeval oproept. De tekst bevat ook emotioneel taalgebruik zoals "leed besparen" of "anderen beschermen".

Daarnaast is in deze bijlage ook een narratieve dimensie opgenomen. Communicatiecampagnes rond risicogedrag maken wel vaker gebruik van zulke narratieve of verhalende boodschappen om gedrag te beïnvloeden (de Graaf et al., 2016; Hinyard & Kreuter, 2007). Deze aanpak is geïnspireerd op de insteek van de algemene preventiecampagne "Een beetje te snel is ook te snel" die in het najaar van 2020 over heel

Vlaanderen en via uiteenlopende kanalen plaatsvond. Die campagne bevatte een waargebeurd verhaal van een autobestuurder die (iets) te snel reed en hierdoor een dodelijk ongeval met een fietser had veroorzaakt. Het verhaal werd een maand lang verspreid via social media filmpjes en radiospots. Op vraag van VSV werd het narratief ook opgenomen in de bijlage. De emotionele brief bevatte dus zowel een visuele component (met de afbeelding van de fiets en het in de verf zetten van de mortaliteitsratio) als een narratief. Beiden zijn onderling met elkaar verbonden aangezien afbeelding en boodschap het verhaal ondersteunen.

3.3 DESIGN EN OPERATIONALISATIE

We onderzochten het effect van de emotionele en de rationele bijlagen bij de boetebrief. Om te achterhalen welke boodschap de intentie om te snel te rijden positief beïnvloedt, maakten we gebruik van een online *randomized survey-experiment* op basis van n = 600 respondenten.

De deelnemers van het experiment werden gerekruteerd via een online panel, dat werd georganiseerd door het marktonderzoekbureau Ipsos naar aanleiding van de evaluatie van de bredere snelheidscampagne. Het survey-experiment werd voorgelegd aan en goedgekeurd door de Sociaal-Maatschappelijke Ethisch Commissie (SMEC) van de KU Leuven[3]. Het experiment liep van 2 april tot 12 april 2021. Deelnemers moesten tussen 18 en 54 jaar oud zijn, in Vlaanderen wonen en in het bezit zijn van een rijbewijs B. Ze werden willekeurig toegewezen aan één van drie mogelijke groepen. Binnen dit randomisatieproces werd gewerkt met quota voor sociaal-demografische parameters (geslacht, leeftijd en provincie) om de steekproef voor Vlaanderen zo representatief mogelijk te maken.

3.3.1 Drie brieven, drie groepen

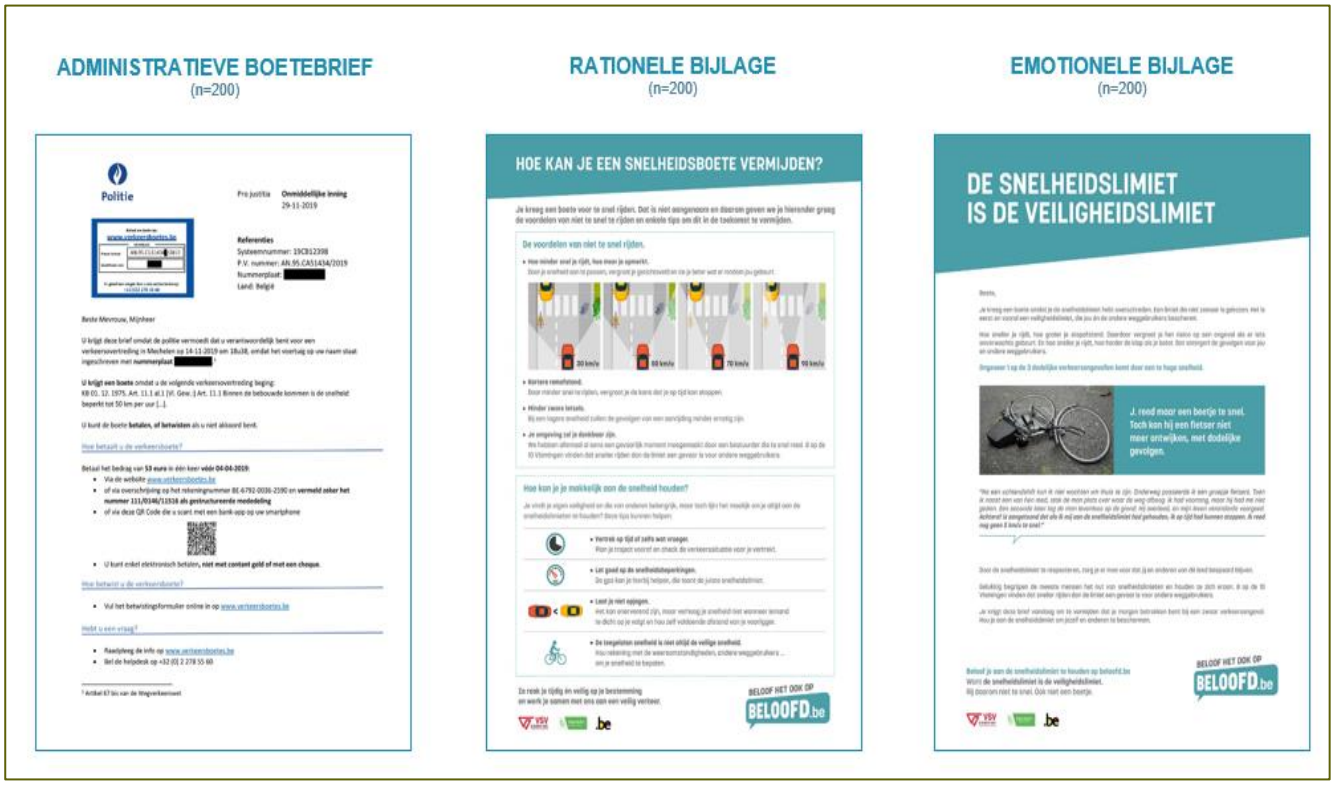
Het experiment heeft betrekking op drie groepen van 200 deelnemers die gevraagd werden een fictieve boetebrief te bekijken. De deelnemers werden opgedeeld in één controlegroep en twee interventiegroepen (zie Tabel 4) die telkens een andere boodschap te zien kregen (zie Figuur 11).

[3] dossier G-2020-1575, datum goedkeuring: 24/11/2020

Tabel 4. Opmaak van groepen in het randomized survey-experiment

GROEP	DEELNEMERS	BRIEF EN BIJLAGE
1	n = 200	Administratieve boetebrief (controle)
2	n = 200	Administratieve boetebrief + rationele bijlage
3	n = 200	Administratieve boetebrief + emotionele/narratieve bijlage

De controlegroep (n=200) werd een vereenvoudigde versie van een reguliere boetebrief voor een snelheidsovertreding getoond. Deze administratieve boetebrief bevatte de details van een fictieve overtreding (tijdstip, plaats, snelheid en nummerplaat), het boetebedrag, de praktische modaliteiten van betaling of van eventuele betwisting. De tweede groep (n=200) kreeg in eerste instantie diezelfde administratieve boetebrief te zien, gevolgd door de rationele bijlage. Deze rationele boodschap werd reeds in de praktijk gebruikt bij het beboeten van snelheidsovertreders van oktober 2019 tot december 2020. De derde groep (n=200) kreeg de administratieve boetebrief voorgelegd in combinatie met de emotionele/narratieve bijlage, die specifiek ontworpen werd voor de snelheidscampagne van 2020.



Figuur 11. Boetebrief met rationele en emotionele bijlage voor het randomized survey-experiment

3.3.2 Intentievragen en PMT constructen

Het experiment bestond uit een voormeting en een nameting. Alle groepen van deelnemers moesten zowel voor als na het zien van de brieven (en bijlage) een aantal intentievragen beantwoorden. Een eerste vraag in de voormeting en in de nameting peilde naar algemene intentie: "Hoe groot acht u de kans dat u zich in de nabije toekomst aan de snelheidslimiet zal houden?".

De vraag werd gevolgd door vijf bijkomende vragen over de intentie om zich aan de snelheidslimiet te houden binnen vijf verschillende snelheidsregimes:

- in een schoolomgeving (30km/u)
- in een woonwijk (30km/u)
- in de bebouwde kom (50km/u)
- buiten de bebouwde kom (70km/u)
- op de snelweg (120km/u)

Voor deze vragen in de voormeting en in de nameting werd gebruik gemaakt van een zevenpunt Likertschaal, die varieerde van een *zeer kleine kans* (1) tot een *zeer grote kans* (7) om zich in de nabije toekomst aan de snelheidslimiet te houden.

Naast de vragen in verband met intentie werd er op basis van de *Protection Motivation Theory* (PMT) ook gekeken naar de onderliggende mechanismen die ons meer zouden kunnen vertellen over waarom de intentie wijzigt (Glendon & Walker, 2013; Kergoat et al., 2017). Deelnemers moesten in de voormeting en in de nameting aangeven in welke mate ze akkoord gingen met zes stellingen gescoord op een schaal van 1 (helemaal niet akkoord) tot 7 (helemaal akkoord). De stellingen werden gebruikt om zes PMT constructen te operationaliseren op basis van de vragenlijst van Glendon & Walker (2013):

- Potentiële kosten (*Potential costs*): "Wanneer ik me nauwgezet aan de snelheidslimieten hou, heb ik het gevoel dat ik tijd verlies"
- Gepercipieerde beloning (*Perceived rewards*): "Ik vind het leuk om sneller te rijden dan de snelheidslimiet"
- Gepercipieerde ernst (*Perceived severity*) "Wanneer ik sneller rijd dan de snelheidslimiet, zal een ongeval erger zijn, dan wanneer ik me aan de snelheidslimiet hou."

- Gepercipieerde kwetsbaarheid: (*Perceived vulnerability*): "Ik heb veel meer kans op een ongeval wanneer ik sneller rijd dan de snelheidslimiet."
- Zelfeffectiviteit (*Self-efficacy*): "Het is voor mij eenvoudig om mij aan de snelheidslimieten te houden."
- Responseffectiviteit (*Response efficacy*): "Door me aan de snelheidslimieten te houden, vergroot de kans dat ik op tijd kan reageren of stoppen bij een onverwachtse situatie."

Tot slot werden enkel in de nameting nog een aantal vragen gesteld over de duidelijkheid, overtuigingskracht, geloofwaardigheid, herkenbaarheid en toon van de rationele en de emotionele boodschappen. Deze vragen werden enkel voorgelegd aan de twee interventiegroepen en niet aan de controlegroep, en gescoord op een schaal van 1 (helemaal niet akkoord) tot 5 (helemaal akkoord). We lichten hierna de resultaten van het experiment in meer detail toe.

3.4 DATA-ANALYSE EN RESULTATEN

De resultaten van ons onderzoek wijzen op het algemene positieve effect van persuasieve boodschappen op de intentie om zich aan de rijsnelheid te houden. De rationele boodschap in ons onderzoek toont meer potentieel dan de emotionele boodschap.

Hieronder volgt een toelichting bij de analyse van de resultaten. In eerste instantie lichten we toe in welke mate het experiment bestond uit een goed uitgebalanceerde groep van deelnemers, ten tweede worden de resultaten voor de algemene intentievraag besproken. Vervolgens wordt het effect van de interventie op de intentie per snelheidsregime toegelicht, en worden de parameters uit de *Protection Motivation Theory* geanalyseerd. We bekijken daarnaast de algemene appreciatievragen over de twee boodschappen, en lichten genderverschillen toe wanneer deze konden worden vastgesteld. Tot slot staan we stil bij een experiment dat vooraf is gegaan aan de resultaten die we hier bespreken, en lijsten we enkele lessen op die we uit dat pre-experiment hebben kunnen trekken.

3.4.1 Balansverhoudingen

Allereerst gaan we na in hoeverre de drie groepen equivalent zijn op de belangrijkste covariaten zoals geslacht, leeftijd en woonplaats. Ook voor andere parameters zoals rijervaring, rijfrequentie en het bezit van een bedrijfswagen gaan we na of de drie groepen statistisch gelijk zijn[4] (zie Tabel 5).

Over de hele steekproef heen bekeken, is de gemiddelde leeftijd 36 jaar. De respondenten in de drie groepen zijn gemiddeld gezien even oud. De verschillen zijn statistisch niet significant (F(2,597) = 0.23, p=0.794). De steekproef bevat ook exact evenveel mannen als vrouwen, een verhouding die ook terug te vinden is in de controlegroep en de twee interventiegroepen afzonderlijk. Iets meer dan de helft van alle deelnemers woont in landelijk gebied (53,8%) terwijl de overige deelnemers in stedelijk gebied (46,2%) wonen. Het aantal deelnemers uit landelijke gebieden, zo toont Tabel 5, is eveneens gelijkmatig verdeeld over de verschillende groepen. De verschillen zijn statistisch niet significant (F(2,597) = 0.41, p=0.665).

Tabel 5. Balance checks: verhoudingen tussen de groepen op basis van enkele parameters

	CONTROLEBRIEF n=200	RATIONELE BIJLAGE n=200	EMOTIONELE BIJLAGE n=200
Mannen (%)	50%	50%	50%
Leeftijd (gem.)	36.50j.	35.80j.	36.39j.
Landelijk (%)	56%	52%	53%
Bedrijfswagen (%)	39,3%	36%	32,7%
Boete (%)	26,9%	23,4%	22,6%
Intentie voormeting (gem.)	5.71	5.69	5.47

Meer dan één op drie deelnemers heeft een bedrijfswagen (36%). Ook hier is dat aantal gelijkaardig voor de drie groepen in het experiment. De relatieve verschillen in Tabel 5 zijn statistisch niet significant (F(2,597) = 0.92, p=0.400). Een kwart van alle deelnemers heeft het voorbije jaar een boete voor te snel rijden in de bus gekregen (24,3%). Ook hier is een gelijke verdeling over de drie groepen terug te vinden van deelnemers die reeds een snelheidsovertreding begingen en zijn de verschillen niet significant (F(2,597) = 0.88, p=0.414). De vergelijkende analyse over deze covariaten bevestigt dus dat de drie groepen mooi in balans zijn – ook

[4] We maken daarbij gebruik van de ANOVA (F-test).

wat betreft een fundamentele parameter, met name de intentie uit de *voormeting* om zich in de nabije toekomst aan de snelheidslimiet te houden (F(2,597) = 1.75, p=0.175).

Gezien het belang van geslacht en leeftijd in verkeersveiligheidsonderzoek in het algemeen en in studies rond rijsnelheid in het bijzonder, loont het de moeite om ook een kruisvergelijking van beiden te maken. De uitsplitsing van de drie groepen in het experiment op basis van geslacht en leeftijdscategorie (zie Tabel 6) toont opnieuw een gelijke verdeling. Over de gehele steekproef zijn er met name 35% deelnemers tussen 18 en 29 jaar, 25% deelnemers tussen 30 en 39 jaar, en 40% deelnemers tussen 40 en 54 jaar. Deze leeftijdsverdeling is binnen dezelfde verhoudingen terug te vinden in de drie groepen in het onderzoek, en vervolgens binnen elke groep bij zowel de mannelijke als de vrouwelijke deelnemers.

Tabel 6. Verhoudingen tussen mannen en vrouwen in de drie groepen volgens hun leeftijd

	CONTROLEBRIEF		RATIONELE BIJLAGE		EMOTIONELE BIJLAGE	
	Man (n=100)	Vrouw (n=100)	Man (n=100)	Vrouw (n=100)	Man (n=100)	Vrouw (n=100)
18-29 jaar	35%	35%	35%	35%	35%	35%
30-39 jaar	25%	25%	25%	25%	25%	25%
40-54 jaar	40%	40%	40%	40%	40%	40%

3.4.2 Effect op de intentie om zich aan de snelheidslimiet te houden

We gaan over naar de hoofdvraag van dit onderzoek: Wat is het effect van de verschillende boodschappen op de intentie van autobestuurders om zich aan de snelheidslimiet te houden?

We meten de intentie om zich aan de rijsnelheid te houden zowel voor als na dat de respondenten in de drie groepen de brieven te zien hebben gekregen. We voeren een gepaarde t-toets uit om te analyseren of de gemiddelde intentie tussen de voor- en de nameting significant verschilt. Op basis van de bevindingen uit andere onderzoeken verwachten we een positief effect van zowel de rationele boodschap als de emotionele boodschap. We verwachten bij deze interventiegroepen dat de intentie om zich aan de rijsnelheid te houden in de nameting zal *stijgen* ten opzichte van de voormeting. Aangezien de

controlegroep geen bijlage te zien kreeg tijdens het experiment is te verwachten dat de intentie in de voormeting en in de nameting niet verschilt.

We stellen vast dat enkel de rationele bijlage een bescheiden, maar significant positief effect heeft op de algemene intentie om zich aan de rijsnelheid te houden. Deze intentie stijgt met .12 op een schaal van 1-7. De administratieve boete op zichzelf blijkt -zoals verwacht- onvoldoende om de intentie van autobestuurders positief te beïnvloeden. Er zijn geen significante verschillen tussen de gemiddelde waarden voor intentie tijdens de voormeting en de nameting. Bovendien heeft de emotionele bijlage geen significant positief effect op de intentie om zich aan de rijsnelheid te houden. Ook hier is er geen significant verschil tussen de voor- en de nameting.

Tabel 7. Intentie algemeen, vergeleken in de voormeting en de nameting[5]

	INTENTIE (gem.)		VERSCHIL (absoluut)	t-WAARDE met t(199)	COHEN'S d (effectgrootte)
	VOOR	NA			
Administratieve Controlebrief	5.71	5.66	-0.05	-0.601	-0.047
Rationele boodschap	5.69	5.81	0.12*	2.158	0.153
Emotionele boodschap	5.47	5.50	0.04	0.359	0.038

Bovenstaand eerste deel van de analyse heeft betrekking op de algemene intentievraag om zich te houden aan de snelheidslimiet. We vroegen de deelnemers in de drie groepen ook naar hun intentie om de snelheidslimiet te respecteren onder verschillende snelheidsregimes. We gaan de effecten van de interventie per regime na in onze analyse en komen tot volgende vaststellingen:

De rationele brief heeft, ten eerste, nog steeds een positief significant effect op intentie in verschillende snelheidsregimes. In de nameting geven respondenten na het zien van de rationele boodschap significant meer aan dat ze zich in de nabije toekomst aan de snelheidslimieten zullen houden – en dat in zowel *schoolomgevingen* (t(199) = 2.387, p=0.009) en *woonwijken* (t(199) = 4.125, p<0.001), als *buiten* de bebouwde

[5] Q"Hoe groot acht u de kans dat u zich in de nabije toekomst aan de snelheidslimiet zal houden?"
* t-test met * voor p < .05, ** voor p < .01, *** voor p < .001

kom (t(199) = 2.484, p=0.007) en op *snelwegen* (t(199) = 2.031, p=0.022). Enkel voor de bebouwde kom vonden we geen significant effect (t(199) = 1.214, p=0.113).

Ten tweede stellen we vast dat ook de emotionele bijlage een positief effect genereert op de intentie zich aan de rijsnelheid te houden. Dit resultaat is verschillend van de algemene intentievraag waar geen significant effect werd gevonden van de emotionele brief. We stellen nu meer bepaald vast dat in de regimes met zone 30 (d.i. in een schoolomgeving en in een woonwijk) het effect van de emotionele bijlage ongeveer even groot als het effect van de rationele bijlage. Dat effect is bovendien even significant. De emotionele bijlage heeft met andere woorden een even positief effect op de intentie van bestuurders om zich aan de snelheid te houden nabij schoolomgevingen (t(199) = 2.384, p=0.009) en woonwijken (t(199) = 3.947, p<0.001) als de rationele brief.

Tabel 8. Intentie in voormeting en nameting weergegeven per groep en per snelheidsregime[6]

	CONTROLEBRIEF (n=200)			RATIONELE BIJLAGE (n=200)			EMOTIONELE BIJLAGE (n=200)		
	VOOR	NA	Verschil (Cohen's d)	VOOR	NA	Verschil (Cohen's d)	VOOR	NA	Verschil (Cohen's d)
Schoolomgeving: Zone 30 km/u	5.88	5.83	-0.05 (-0.051)	5.90	6.06	0.16** (0.169)	5.84	5.99	0.15** (0.168)
Woonwijk: Zone 30 km/u	5.28	5.51	0.23** (0.175)	5.42	5.73	0.31*** (0.294)	5.28	5.57	0.29*** (0.288)
Binnen bebouwde kom: 50 km/u	5.82	5.82	-0.01 (0.006)	6.00	6.08	0.08 (0.086)	5.71	5.79	0.08 (0.088)
Buiten bebouwde kom: 70 km/u	5.57	5.47	-0.10 (0.105)	5.56	5.76	0.19** (0.178)	5.40	5.48	0.09 (0.102)
Snelweg: 120 km/u	5.54	5.54	0.00 (0.004)	5.60	5.75	0.14* (0.143)	5.37	5.41	0.04 (0.045)

Ten derde vinden we dat de administratieve boetebrief alleen een bescheiden effect sorteert als het gaat om woonwijken met zone 30 (t(199) = 2.438, p=0.008). Het effect is statistisch significant maar iets kleiner in vergelijking met het effect dat gegenereerd werd in de interventiegroepen met de rationele en emotionele

[6] Q: Hoe groot acht u de kans dat u zich in de nabije toekomst aan de snelheidslimiet zal houden in …?
t-test met * voor p < 0,05, ** voor p < 0,01, *** voor p < 0,001

bijlagen. Het valt, ten vierde, ook op dat de intentie om zich binnen de bebouwde kom aan de rijsnelheid te houden niet significant wijzigt tijdens de voor- en de nametingen in de drie groepen.

Het is, ten slotte, interessant om op enkele genderverschillen te wijzen die uit de resultaten naar voor komen. Zo heeft de rationele brief een positief effect bij zowel mannen ($t(99) = 2.568$, $p=0.006$) als vrouwen ($t(99) = 3.543$, $p<0.001$) op hun intentie zich aan de rijsnelheid te houden in een woonwijkregime 30km/u. De significantieniveaus liggen wel hoger bij vrouwen dan bij mannen. Belangrijkere verschillen tussen mannen en vrouwen voor het effect van de rationele brief worden vastgesteld bij de regimes in een *schoolzone* en in de zones *buiten* de bebouwde kom (70km/u en 120km/u): In de schoolzone is er *wel* een statistisch significant en positief effect van de rationele brief op de intentie bij vrouwen ($t(99) = 1.968$, $p=0.026$) maar *niet* bij mannen ($t(99) = 1.495$, $p=0.069$). Daarentegen heeft de rationele brief in de regimes buiten de bebouwde kom *geen* significant effect heeft bij vrouwen, en *wel* bij mannen (voor het regime 70km/u geeft de teststatistiek $t(99) = 2.241$, $p=0.014$ en voor het regime 120km/u is dat $t(99) = 1.784$, $p=0.039$). Gelijkaardige vaststellingen kunnen worden gemaakt voor de emotionele brief. Ook hier genereert de boodschap een positief effect bij mannen en bij vrouwen in een woonwijkregime 30km/u (met de resultaten van de teststatistiek op resp. $t(99) = 2.666$, $p=0.005$ en $t(99) = 2.910$, $p=0.002$). Bovendien zijn enkel vrouwen positief en significant beïnvloed door de emotionele boodschap in een schoolzone ($t(99) = 2.623$, $p=0.005$). Buiten de bebouwde kom, in het regime 70km/u, is er dan weer *wel* een positief significant effect van de emotionele boodschap op mannen ($t(99) = 1.827$, $p=0.035$) maar *niet* op vrouwen ($t(99) = 0.332$, $p=0.370$).

Uit dit experiment lijkt te volgen dat mannen meer geneigd zijn hun gedrag aan te passen in de zones buiten de bebouwde kom na het lezen van beide boetebrieven, terwijl vrouwen gevoeliger zijn voor de rationele en emotionele boodschap wanneer het een schoolzone betreft.

3.4.3 Effect op de drijfveren van verandering in intentie via de Protection Motivation Theory

Onderzoeken of een bepaalde boodschap en een bepaalde communicatiestijl een effect hebben, is een waardevol onderdeel van beleidsevaluatie. In de vorige stap gingen we na *wat* werkt. Een volgende stap is trachten te achterhalen *waarom* het werkt.

Om het evaluatieproces te verrijken en de drijfveren van een verandering in intentie te doorgronden, baseren we ons op de *Protection Motivation Theory*. Tabel 9 geeft de PMT items weer en de gemiddelde scores op deze items op een schaal van 1 tot 7 tijdens de voormeting en de nameting voor de respondenten in de controlegroep en de twee interventiegroepen. Ook hier toont de analyse aan dat de rationele boodschap een statistisch significant effect ressorteert. Drie van de zes PMT constructen worden significant beïnvloed.

Tabel 9. Effecten van de boodschappen op de PMT constructen[7]

	CONTROLEBRIEF (n=200)			RATIONELE BIJLAGE (n=200)			EMOTIONELE BIJLAGE (n=200)		
	VOOR	NA	Verschil (Cohen's d)	VOOR	NA	Verschil (Cohen's d)	VOOR	NA	Verschil (Cohen's d)
Costs	3.91	3.94	0.03 (0.028)	3.96	3.88	-0.08 (-0.089)	3.89	3.96	0.07 (0.059)
Rewards	3.55	3.43	-0.13 (0.102)	3.62	3.46	-0.17* (-0.132)	3.65	3.75	0.10 (0.073)
Perceived severity	5.64	5.66	0.02 (0.017)	5.71	5.74	0.04 (0.036)	5.54	5.51	-0.03 (-0.032)
Perceived vulnerability	5.34	5.37	0.02 (0.022)	5.21	5.44	0.22*** (0.230)	5.03	5.28	0.25*** (0.258)
Self-efficacy	5.03	5.14	0.11* (0.129)	5.23	5.45	0.21*** (0.270)	4.89	4.99	0.09 (0.088)
Response efficacy	5.82	5.68	-0.14* (-0.139)	5.87	5.95	0.09 (0.113)	5.67	5.73	0.06 (0.068)

Concreet slaagt de rationele boodschap erin om de perceptie van kwetsbaarheid te verhogen (*perceived vulnerability)* of de mate waarin men gelooft dat de kans op een ongeval groter wordt door te snel te rijden. Ook het gevoel van beloning (*rewards*), waarbij te snel rijden als iets aangenaams wordt ervaren, vermindert significant in de nameting ten opzichte van de voormeting. En tot slot slaagt de rationele brief er ook in om bestuurders significant meer het gevoel te geven dat ze zich eenvoudiger aan de snelheidslimieten zullen kunnen houden. De zelfeffectiviteit (*self-efficacy*) neemt toe. De tips rond het respecteren van de rijsnelheid in de rationele bijlage kunnen een verklaring bieden voor de resultaten rond

[7] Q: In welke mate bent u akkoord met volgende stellingen?
t-test met * voor p < 0,05, ** voor p < 0,01, *** voor p < 0,001

deze PMT constructen. De rationele brief lijkt immers zowel het dreigingsproces als het coping proces binnen het PMT model te capteren.

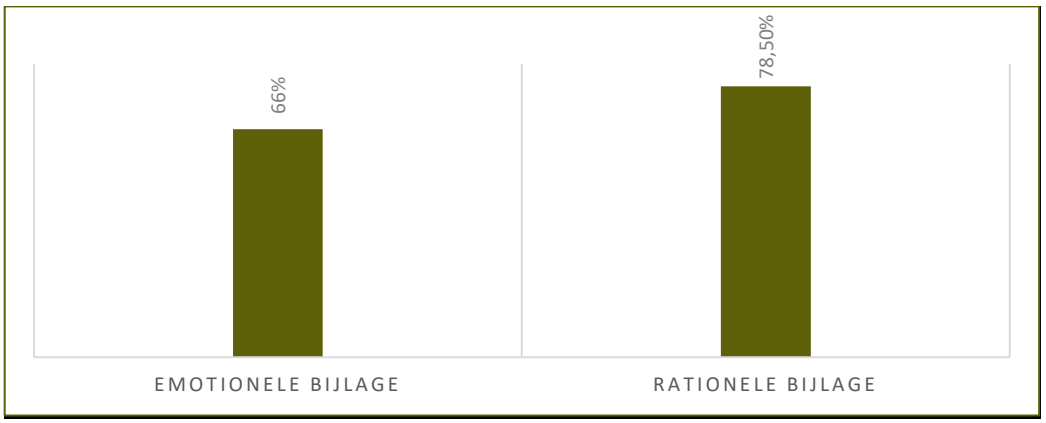
De emotionele brief daarentegen haalt dan wel het hoogste significante effect op de perceptie van kwetsbaarheid (*perceived vulnerability*), maar de emotionele boodschap slaagt er onvoldoende in om mensen ook antwoorden en oplossingen aan te reiken. Het coping proces wordt door deze brief niet voldoende gecapteerd. De brief bevat bijvoorbeeld de zin "door de snelheidslimiet te respecteren, zorg je er mee voor dat jij en anderen van dit leed bespaard blijven," maar biedt verder geen specifieke tips of instructies om dit ook te realiseren.

Bij de boetebrief, ten slotte, merken we een lichte stijging op het vlak van zelfeffectiviteit (*self efficacy*). Deze stijging is kleiner in vergelijking met het effect van de rationele brief, en statistisch ook zwakker. Bovendien wordt dit effect onmiddellijk tenietgedaan door een daling inzake responseffectiviteit (*response efficacy*) tijdens de nameting in vergelijking met de voormeting. De boetebrief in de controlegroep leidt met andere woorden tot een lagere inschatting van de effectiviteit van aangereikte antwoorden om het gewenst gedrag te vertonen en zo de dreiging af te wenden. Verder tonen de resultaten dat de boetebrief alleen niet volstaat om de perceptie van dreiging *(perceived vulnerability*) te versterken.

3.4.4 Appreciatie van de rationele en emotionele boodschap

We evalueerden in ons onderzoek de algemene appreciatie van de rationele versus de emotionele brief. De appreciatievragen werden logischerwijs enkel gesteld na het zien van de bijlagen, en dus is er hier geen voormeting. Ook de groep deelnemers die enkel de controlebrief te zien kreeg, wordt hier niet vermeld.

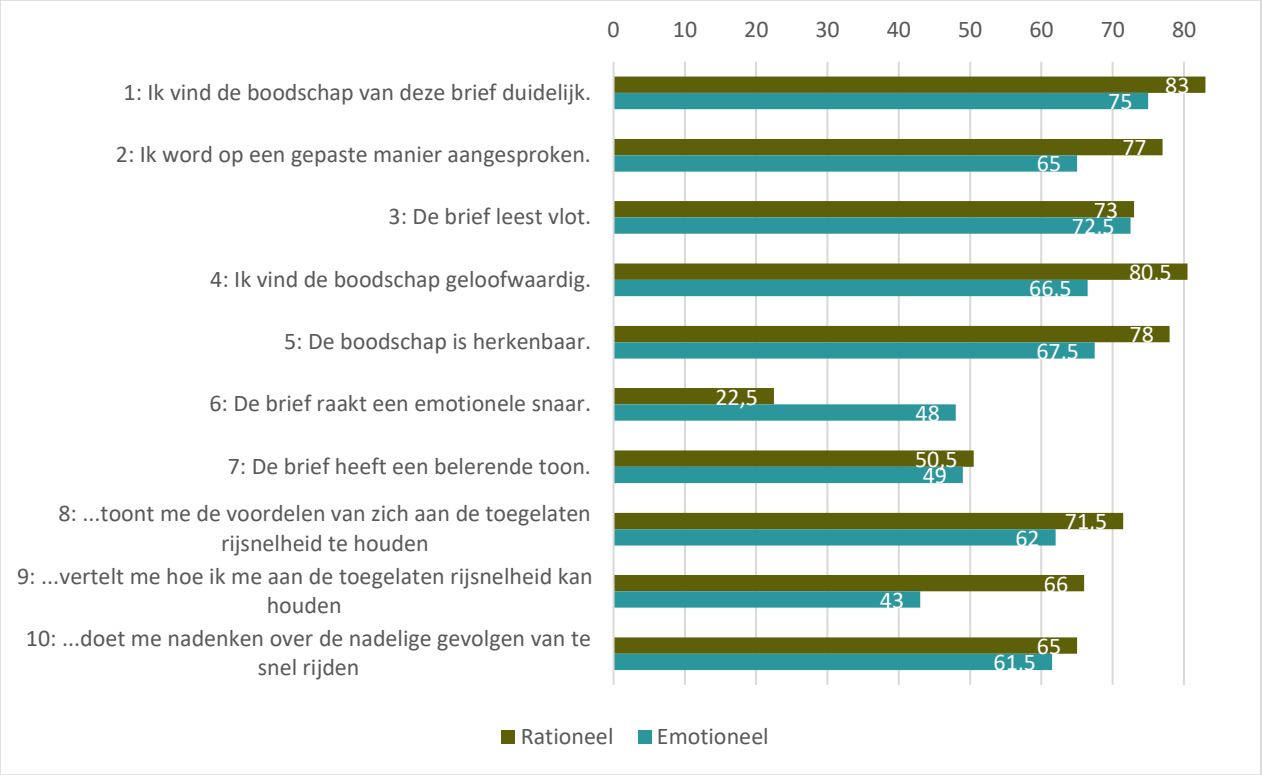
Figuur 12 geeft het relatieve aantal respondenten weer dat de brief een algemene appreciatiescore van 7 of meer op 10 gaf. Er is een uitgesproken verschil tussen de appreciatiescores van de rationele brief en de emotionele brief. De rationele bijlage werd namelijk door bijna 80% van de deelnemers in de groep gescoord met 7/10 of meer, terwijl deze score gegeven werd door twee derde van de deelnemers die de emotionele brief te zien kregen. Het algemene appreciatiegevoel is dus hoger bij de deelnemers die de rationele brief te zien kregen.



Figuur 12. Aandeel deelnemers met appreciatiescore 7/10 of meer in de twee interventiegroepen

Op basis van de gedetailleerde appreciatievragen rond duidelijkheid, vlotheid, geloofwaardigheid etc. stellen we vast dat de emotionele brief minder duidelijk, geloofwaardig en herkenbaar overkomt voor deelnemers aan het experiment. De scores voor de emotionele brief zijn relatief gezien nog vrij hoog, maar ze liggen lager dan bij de rationele brief. Bovendien geven ook minder deelnemers die de emotionele brief te zien kregen aan dat deze brief hen vertelt hoe ze zich dan wél aan de toegelaten rijsnelheid kunnen houden. Slechts 62% van de deelnemers gaat ermee akkoord dat de emotionele brief dit doet, tegenover 71,5% van de deelnemers die de rationele brief te lezen kreeg. Hiertegenover staat dat de emotionele brief er wel bij veel meer deelnemers in slaagt een emotionele snaar te raken. Bijna de helft van de deelnemers in de groep die de emotionele bijlage las, gaat hiermee akkoord ten opzichte van slechts 22,5% van de deelnemers die de rationele brief voorgelegd kregen. Het emotionele karakter van de brief vertaalt zich echter niet naar een hogere algemene appreciatie.

We kunnen besluiten dat de rationele brief over de gehele lijn beter uit de verf komt dan de emotionele brief. Figuur 13 toont een overzicht van de gedetailleerde appreciatievragen en het relatieve aantal deelnemers dat ermee akkoord ging in de twee interventiegroepen.



Figuur 13. Percentage deelnemers dat 'helemaal' of 'eerder' akkoord gaat met de appreciatievragen

Er zijn enkele genderverschillen vast te stellen wat betreft de mate waarin de twee brieven worden geapprecieerd[8]. De gemiddelde score die vrouwen op een schaal van 1 (helemaal niet akkoord) tot 5 (helemaal akkoord) toekennen aan de appreciatievragen ligt over het algemeen hoger dan die van mannen, zowel wat de rationele brief betreft (3.82 vs 3.65) als wat de emotionele brief aangaat (3.72 vs. 3.60). Bovendien wordt de rationele brief door vrouwen statistisch significant hoger ingeschat dan door mannen op de items 'duidelijkheid', 'gepaste aanspreking', 'vlotheid' en 'geloofwaardigheid'. De emotionele brief daarentegen toont een significant hogere appreciatie van vrouwen op de items 'geloofwaardigheid' en 'nadenken over nadelige gevolgen'.

[8] Op basis van gepaarde t-testen

3.4.5 Lessen leren en repliceren: het pre(lude)-experiment

Het survey-experiment dat uitvoerig werd toegelicht betreft in feite een vervolgexperiment. Het is een quasi replicatie van een eerste experiment dat werd uitgevoerd in december 2020. De resultaten van dit eerste experiment (met significante effecten voor de rationele, maar ook voor de emotionele brief) kunnen we echter niet meenemen in de finale rapportage. De reden hiervoor is dat het randomisatieproces toen niet correct verlopen is. Het proces zelf levert echter wel enkele belangrijke lessen op die in het kader van een rapport over het opzetten van experimenten in beleid niet mogen ontbreken. We bespreken kort de drie belangrijkste lessen:

Les 1: Zorg dat de randomisatie goed zit en de groepen gelijk verdeeld zijn

Op het eerste gezicht leken de groepen bij het eerste experiment gelijk verdeeld op vlak van leeftijd en geslacht. Een diepere analyse en kruisvergelijking op basis van de leeftijdscategorieën leverde toen echter een heel ander beeld op (zie Tabel 10). Uit de kruisvergelijking bleek dat jongere mannen fors ondervertegenwoordigd waren in de drie groepen in het eerste experiment. Ook het aandeel jonge vrouwen lag erg hoog ten opzichte van de vrouwen in de andere leeftijdscategorieën. De groepen waren dus niet equivalent waardoor ook de resultaten vertekend waren.

Tabel 10. Verdeling groepen naar geslacht en leeftijd in pre-experiment

	CONTROLEBRIEF		RATIONELE BIJLAGE		EMOTIONELE BIJLAGE	
	Man (n=92)	Vrouw (n=107)	Man (n=107)	Vrouw (n=94)	Man (n=99)	Vrouw (n=101)
18-29 jaar	8%	59%	16%	51%	11%	59%
30-39 jaar	32%	15%	34%	24%	29%	20%
40-54 jaar	61%	26%	50%	24%	60%	21%

Les 2: De formulering (framing) van de intentievraag is belangrijk

Bij het eerste experiment hebben we de algemene vraag naar intentie geformuleerd vanuit het *ongewenste* gedrag, met name te snel rijden. De exacte vraag luidde toen als volgt: "Hoe groot acht u de kans dat u in de nabije toekomst *sneller zou rijden* dan de snelheidslimiet?" In totaal gaf gemiddeld 23,5% van de 600 respondenten bij de nameting aan dat de kans eerder groot, groot of zeer groot was dat ze in de nabije toekomst te snel zouden rijden. Bij het tweede experiment hebben we de vraagstelling omgedraaid en geformuleerd vanuit het *gewenste* gedrag omdat dit beter de rationele van de *Protection Motivation Theory*

weergeeft. De PMT draait immers om de motivatie om zich te beschermen. De vraag luidde nu: "Hoe groot acht u de kans dat u zich in de nabije toekomst *aan de snelheidslimiet zal houden*? Slechts 7% van de 600 respondenten gaf nu bij de nameting aan dat de kans eerder klein, klein of zeer klein was dat ze zich niet aan de snelheidslimieten zouden houden. Mensen ervaren de vraag naar de intentie om te snel te rijden (ongewenste gedrag) en de intentie om zich aan de snelheidsregels te houden (gewenst gedrag) duidelijk anders.

Les 3: Heb aandacht voor onverwachte persaandacht

Terwijl we nog volop bezig waren met de analyse van het eerste experiment werden een aantal voorlopige resultaten gepresenteerd op het Vlaamse Congres Verkeersveiligheid van maart 2021. Deze resultaten werden meteen opgepikt door de pers. In en door de pers werd een stijging van intentie niet weergegeven in procent*punten* maar in procenten, waardoor de resultaten spectaculairder oogden dan ze werkelijk waren[9]. Op die manier werd het artikel ook door andere media snel opgepikt. Het lijkt aangewezen om de communicatie over de resultaten van wetenschappelijk onderzoek zelf goed in de hand te houden. Dit kan door vooraf een uitgekiend communicatieplan en duidelijke afspraken tussen beleidsmakers en onderzoekers.

3.5 CONCLUSIES

Uit de evaluatie van rationele versus emotionele boodschappen in dit gerandomiseerde survey-experiment blijkt dat de rationele boodschap op alle vlakken beter scoort. De rationele bijlage die aan de boetebrief werd toegevoegd heeft een significant effect op algemene intentie om zich aan de rijsnelheid te houden, werkt binnen vier van de vijf snelheidsregimes, en slaagt erin om drie van de zes factoren van de PMT gunstig te beïnvloeden. In vergelijking met de emotionele bijlage in dit experiment krijgt de brief ook een hogere algemene appreciatie en scoort hij beter op het vlak van duidelijkheid, geloofwaardigheid en herkenbaarheid. Tot slot toont de rationele boodschap ook relatief goed aan hoe men zich wel aan toegelaten rijsnelheid kan houden.

[9] Voor persaandacht zie bijvoorbeeld: https://www.nieuwsblad.be/cnt/dmf20210326_98101813; https://www.hln.be/binnenland/snelheidsboetes-komen-voortaan-met-brief-van-doodrijder~a15f73bf/

De boetebrief en de emotionele brief slagen er niet in om de intentie om zich aan de rijsnelheid te houden significant te verbeteren. Toch heeft de emotionele brief significant positieve effecten in schoolomgevingen en woonwijken (beide zone 30), raakt hij overduidelijk een gevoelige snaar en doet hij ook de perceptie van kwetsbaarheid significant toenemen. Verder mogen we concluderen dat het versturen van een administratieve boete op zichzelf geen of nauwelijks effect heeft op het rij-intentie van autobestuurders.

In tegenstelling tot de beperkte maar significante effecten bij het veldexperiment in het Verenigd Koninkrijk vinden we in dit survey-experiment in Vlaanderen dus geen significante effecten terug voor de emotionele brief. In wat volgt bespreken we een aantal mogelijke verklaringen voor dit resultaat.

Een eerste element dat in het oog springt heeft betrekking op het design en het type van experiment. We gebruikten in het Vlaamse onderzoek een survey-experiment en geen veldexperiment zoals in het Verenigd Koninkrijk. Een veldexperiment meet 'werkelijk' gedrag in de 'echte' wereld, terwijl ons survey-experiment enkel keek naar de zelf-gerapporteerde intentie. Hoewel recente studies aantonen dat er een sterke correlatie is tussen de intentie om te snel te rijden en effectief te snel rijden (Elliott et al., 2003; Haglund & Åberg, 2000), kunnen we de *intention-behaviour gap* niet zomaar wegdenken. Intentie en gedrag kunnen met andere woorden niet vanzelfsprekend worden gelijkgeschakeld (Armitage & Conner, 2001). Ten tweede zijn er ook substantiële verschillen tussen V.K. en Vlaanderen op vlak van grootte en samenstelling van de steekproef in het experiment. Het veldexperiment in het V.K. betrof 11.010 deelnemers, waarvan de emotionele bijlage enkel werd getoond aan mensen die een snelheidsovertreding hadden begaan. In Vlaanderen werd met het survey-experiment gestreefd naar 600 respondenten die een representatieve steekproef vormden van alle autobestuurders met een rijbewijs B tussen 18 en 54 jaar.

Tot slot bevat de emotionele bijlage één fundamenteel manco. Een cruciale voorwaarde waaraan niet voldaan is in ons experiment, is dat de boodschap handvatten aanreikt om de dreiging te verminderen. Iets wat overigens ook niet in de brief uit het VK vervat zat. Theorie en empirie naar angstboodschappen toont aan dat zij er erg goed in slagen om de aandacht te trekken, en mensen meer bewust kunnen maken van ongewenst en gewenst gedrag. Dreigen en angst aanjagen volstaan op zichzelf echter niet om gedrag te veranderen (Witte & Allen, 2000). Hiervoor zijn geloofwaardige handelingsperspectieven en concrete tips

nodig (Wundersitz et al., 2010). Vanuit het PMT oogpunt is de *threat appraisal* dus wel sterk aanwezig in onze emotionele boodschap, maar ontbreekt het andere luik met name de *coping appraisal*. In tegenstelling tot de rationele brief geeft de emotionele brief geen tips om je aan de rijsnelheid te houden, en slaagt hij er niet in de zelfeffectiviteit en de responseffectiviteit te verhogen. Met andere woorden, de angstboodschap op zichzelf volstaat niet, maar dient te worden vergezeld van een bepaalde vorm van advies (Elliott et al., 2003). Ook de resultaten rond de apprecitatievragen van de emotionele boodschap wijzen in deze richting. Het narratief is voor de ontvangers minder herkenbaar en geloofwaardig dan de rationele boodschap.

3.6 AANBEVELINGEN

In deze laatste paragraaf formuleren we nog twee aanbevelingen volgend uit de resultaten van het snelheidsexperiment. De eerste aanbeveling heeft betrekking op het design van de boodschap. De tweede aanbeveling heeft betrekking op de evaluatie.

Wat het design van de boodschap betreft, stellen we op basis van dit onderzoek voor om een nieuwe bijlage te ontwerpen die het beste van beide boodschappen combineert en het narratief weglaat. Op die manier kunnen we komen tot een geloofwaardige boodschap die een mix van zowel rationele elementen als emotionele elementen bevat. Onderzoek rond rijgedrag en alcohol heeft reeds aangetoond dat een brief die rationele en emotionele ingrediënten combineert het meeste potentieel heeft (Warner & Forward, 2016). Oorspronkelijk was een gecombineerde brief ook opgenomen in het design van deze studie, maar wegens de beperkte steekproefgrootte en met het oog op de nodige analytische power werd besloten om die 'hybride' brief te laten vallen. Het voordeel van een gecombineerde brief is als volgt: Ten eerste zou de emotionele kant van de brief de aandacht trekken en zo de perceptie van zowel de kans op als de ernst bij een auto-ongeval kunnen verhogen. Ten tweede zou de rationele kant van de brief gewenst gedrag kunnen faciliteren via concrete tips en handelingsperspectieven. Door op deze gecombineerde manier in te spelen op de PMT processen *threat appraisal* en *coping appraisal* verwachten we dat de motivatie van mensen om zich te beschermen en de intentie om zich aan de rijsnelheid te houden nog sterker zou toenemen.

De tweede aanbeveling betreft de evaluatie van snelheidsbrieven en de impact ervan op rijgedrag. Wij zijn overtuigd van een *evidence-based* aanpak die de effecten van interventies op het rijgedrag *op voorhand*

verkent en tracht in te schatten. Een nieuwe brief kan op basis van hetzelfde design via een survey-experiment op voorhand getest en geëvalueerd worden vooraleer deze in de brievenbus van snelheidsovertreders terecht komt. Naast een survey-experiment kunnen beleidsmakers ook overwegen om in een tweede fase een veldexperiment op te zetten met een groep recidivisten als steekproef. Op die manier kan wetenschappelijk onderzoek meer robuust inzicht geven in het feitelijke gedrag van bestuurders die snelheidsovertredingen begaan en zou kunnen worden aangetoond in welke mate intentie zich verhoudt tot het werkelijke gedrag.

4 VELDEXPERIMENT PARKEERBOETES

Sinds 2014 zijn verkeersovertredingen in verband met foutief parkeren of stilstaan vervat in de wetgeving rond gemeentelijke administratieve sancties en behoren ze tot de familie van de GAS-boetes. Op basis van deze GAS-wetgeving zijn lokale besturen niet langer genoodzaakt om dit soort kleinere inbreuken op het politiereglement strafrechtelijk te handhaven, maar kunnen ze zelf via een administratieve sanctie sneller optreden tegen verkeersovertredingen. Burgers storen zich bijvoorbeeld aan wagens die foutief geparkeerd zijn of foutief stilstaan. Deze voertuigen belemmeren de mobiliteit van voetgangers, fietsers, auto's en andere weggebruikers. Daarnaast hebben parkeerovertredingen een negatieve invloed op de veiligheid en de levenskwaliteit in stadscentra en gemeenten (Schram & Lievens, 2015).

Vooral steden hebben op deze manier een extra instrument ter beschikking om een coherent, efficiënt en effectief parkeerbeleid uitbouwen. Tegelijkertijd is het hoge en steeds toenemende aantal vastgestelde verkeersovertredingen uitgegroeid tot een belangrijke inkomstenbron voor grote(re) steden. Volgens cijfers van de Vereniging voor Vlaamse Steden en Gemeenten (VVSG) en het Vlaams Agentschap Binnenlands Bestuur (ABB) inden Vlaamse gemeenten in 2014 € 4,4 miljoen aan GAS-boetes. In 2018 was dit bedrag bijna vertienvoudigd tot € 42,6 miljoen (Verbergt, 2020). Tussen 2015 en 2020 is het aantal GAS-boetes voor parkeerovertredingen zelfs met een factor 13 toegenomen (IBZ, 2021). Het is te verwachten dat deze inkomsten in de toekomst nog verder zullen toenemen door de opmars van slimme ANPR-camera's en de beslissing van de Vlaamse regering om lokale besturen vanaf 1 februari 2021 de mogelijkheid te geven beperkte snelheidsovertredingen via een GAS-boete te handhaven. Momenteel zetten vooral steden zoals Antwerpen, Leuven, Gent, Genk, Oostende en Mechelen hier sterk op in (IBZ, 2021; Schram & Lievens, 2015).

Handhaving via GAS-boetes is geen doel op zich, maar een middel om gemeenten veiliger en leefbaarder te maken. Niettemin is het voor lokale besturen belangrijk dat deze boetes correct en tijdig geïnd worden. Op die manier zijn de inkomsten gegarandeerd en de administratieve (werk)last beperkt. Daarnaast voorkomt het correct en tijdig betalen van deze parkeerboetes ook mogelijke extra kosten voor de burger. In dit hoofdstuk onderzoeken we of en in welke mate subtiele gedragsinterventies zoals nudges, opgenomen in

de boetebrief die verstuurd wordt naar parkeerovertreders, het aantal tijdige betalingen effectief doen toenemen.

Concreet testen we op basis van een veldexperiment (n = 2664) in de Stad Mechelen het effect van drie gedragsinterventies, zijnde het vereenvoudigen van de boetebrief, het toevoegen van een afschrikwekkende boodschap en het toevoegen van een normerende boodschap op basis van descriptieve, sociale normen. We gaan na of deze interventies de betalingsgraad positief kunnen beïnvloeden. Hiervoor werken we samen met GASAM, de interlokale vereniging die instaat voor de administratieve en financiële opvolging van de gemeentelijke administratieve sancties in het arrondissement Mechelen, en met het Team Gedragsinzichten van het Departement Kanselarij en Buitenlandse Zaken (DKBUZA) van de Vlaamse overheid. Het experiment kreeg de ambtelijke en politieke steun van de Stad Mechelen en werd ingediend bij en goedgekeurd door de Sociaal-Maatschappelijke Ethische Commissie (SMEC) van de KU Leuven.[10]

Voor het theoretische kader en het operationaliseren van de interventies inspireerden we ons op de uitgebreide literatuur rond nudging in een belastingcontext. Qua design maken we gebruik van een Cluster Randomized Controlled Trial met een controlegroep en drie interventiegroepen. De keuze voor een geclusterd design is ingegeven door de procedure die GASAM hanteert voor het uitsturen van de boetebrieven, namelijk op basis van batches van opgespaarde boetes per aantal dagen. In dit hoofdstuk benaderen we nudging niet als een afzonderlijk beleidsinstrument tegen overlast of ter bevordering van verkeersveiligheid, maar onderzoeken we hoe we via nudging een onderdeel van een bestaand beleidsinstrument (de betaling van een GAS-boete voor foutief stilstaan of parkeren) kunnen versterken. In Box 3 geven we als samenvatting de belangrijkste elementen van deze case mee op basis van het stappenplan voor het opstellen van een gedragsinterventie, zoals geschetst in het eerste hoofdstuk.

[10] Dossiernummer G-2020-1435, datum goedkeuring: 22/05/2020

Stappenplan voor het opzetten van gedragsinterventies: case parkeerboetes

Stap 1: Definitiefase

Wat is het probleemgedrag? Tussen de 30% en 35% van de personen die een GAS-boete krijgen voor fout stilstaan of parkeren in de Stad Mechelen betalen deze boete te laat. Wat is het gewenste gedrag? Personen die een GAS-boete krijgen voor fout stilstaan of parkeren in de Stad Mechelen betalen deze op tijd. Wie is de doelgroep? Alle personen die een GAS-boete krijgen voor fout stilstaan of parkeren in de Stad Mechelen, bedrijfswagens niet meegerekend.

Stap 2: Analysefase

Wat zijn de determinanten om te laat te betalen? We baseren ons op literatuur inzake belastingen en de checklist voor brieven en mails van de Vlaamse Overheid en het *Behavioural Insights Netwerk Nederland* (BIN NL 2018) om te onderzoeken of de brief de informatie goed overbrengt, de aandacht trekt, motiverend werkt, weerstand tegengaat en de lezer aanzet tot actie. We concluderen dat we op deze vijf factoren zullen inspelen: lay-out + toon, aandacht, motivatie, weerstand en uitstelgedrag.

Stap 3: De interventiefase

Op basis van de literatuur rond nudging in een belastingcontext, in combinatie met de vier principes van het EAST framework, selecteren we drie interventies:

1. we maken de boodschap gemakkelijk en vereenvoudigen de brief
2. we maken de boodschap opvallend en accentueren mogelijke extra kosten
3. we maken de boodschap sociaal en beschrijven wat de meeste mensen doen

Stap 4: De designfase

We ontwerpen een *Cluster Randomized Controlled Trial* met een controlegroep en drie interventiegroepen. De keuze voor een geclusterd design is ingegeven door de procedure die GASAM hanteert voor het uitsturen van de boetebrieven, namelijk op basis van batches van opgespaarde boetes per aantal dagen. Elke batch krijgt één van de vier interventies toegewezen.

Stap 5: De evaluatiefase

We zetten het geobserveerde percentage van op tijd betaalde boetes per type brief af tegen het geobserveerde percentage niet op tijd betaalde boetes. Vervolgens testen we de significantie en effectgroottes van de interventiebrieven ten opzichte van de controlebrief door middel van een multi-level logistische regressieanalyse. Deze resultaten vertalen we naar concrete beleidsaanbevelingen.

Box 3. Samenvatting case parkeerboetes volgens stappenplan

4.1 SCOPE: VAN GAS TOT GASAM

In dit eerste luik zullen we dieper ingaan op de eigenheid van de GAS wetgeving inzake fout parkeren en stilstaan en besteden we aandacht aan het specifieke karakter van de GAS procedures binnen de Stad Mechelen, die relevant zijn voor het experiment. Daarbij zoomen we in op de werking van GASAM, de interlokale vereniging die instaat voor de administratieve opvolging van de gemeentelijke administratieve sancties in het arrondissement Mechelen. Binnen deze werking focussen we vooral op de betalingstermijn en de betalingsgraad, Na een overzicht van de procedures werpen we een blik op de belangrijkste cijfers van GASAM. We geven mee hoeveel dossiers jaarlijks worden behandeld, over welke type overtredingen het gaat en wat de huidige verhouding is tussen tijdig betaalde en niet tijdig betaalde parkeerboetes.

4.1.1 GAS boetes voor fout parkeren en stilstaan

Gemeentelijke administratieve sancties vormen een beleidsantwoord op het probleem van openbare overlast (Schram & Lievens, 2015). Het betreft een straf, doorgaans een boete, die lokale besturen kunnen opleggen voor inbreuken op het gemeentelijke politiereglement. Hoewel deze regelementen niet allemaal exact dezelfde elementen omvatten, kunnen we een aantal grote categorieën onderscheiden: openbare rust (bv. nachtlawaai); vervuiling (bv. sluikstorten of wildplassen); beschadiging van goederen (bv. vandalisme, graffiti); verkeer (parkeren) (Schram & Lievens, 2015).

Het huidige systeem van gemeentelijke administratieve sancties bouwt voort op de wet van 13 mei 1999 (Allewaert & De Sutter, 2010). Op basis van de GAS-wetgeving hoeven lokale besturen voor minder ernstige inbreuken die zorgen voor openbare overlast niet langer de weg van de strafrechtelijke handhaving te volgen. Ze krijgen zelf de bevoegdheid en dus de mogelijkheid om dit via een administratieve sanctie te beboeten (Cops et al., 2012). Aangezien het Openbaar Ministerie en de politierechter de prioriteit leggen bij zwaarder criminele inbreuken, blijven kleinere overtredingen nu minder onbestraft. Zo vermijden lokale besturen de perceptie dat dit soort gedrag aanvaardbaar is (De Schepper et al., 2013). Naast het sneller optreden tegen verschillende vormen van overlast, wordt ook het gerechtelijke apparaat deels ontlast via de gemeentelijke administratieve sancties (Schram & Lievens, 2015).

Ondanks de voordelen zijn GAS-boetes hun invoering onderwerp van controverse en kritiek (Devroe, 2012). Vooral als het gaat om overlast waarschuwen critici voor het risico op zogenaamde "*net widening*", een proces waarbij steeds meer aspecten van de samenleving voorwerp worden van strafrechtelijk beleid (Schram & Lievens, 2015). GAS-boetes hebben voor burgers soms een negatieve connotatie omdat ze onduidelijk en arbitrair zouden zijn (Bauwens et al., 2016; Cops et al., 2012). Ook de rigide manier waarop enkele gemeenten deze nieuwe instrumenten in de beginjaren invoerden, versterkte deze perceptie. Er circuleerden termen zoals "Gigantisch Absurde Sancties", "pestboetes" en "verGASsing" (Eeckhout, 2013; Schram & Lievens, 2015). Ook recent nog waren GAS-boetes voor het overtreden van de coronamaatregelen of het rondlopen in zwemoutfit in bepaalde kustgemeenten het mikpunt van kritiek en spot (Langbeen & Thomas, 2020).

Die negatieve perceptie is dus niet helemaal verdwenen. Toch maakt 95% van de Belgische gemeenten vandaag gebruik van de administratieve sancties (IBZ, 2021). De GAS-wetgeving werd ook op geregelde tijdstippen aangepast en hervormd.[11] In 2013 kwamen er naast gemeentelijke administratieve sancties voor overlast ook gemeentelijke administratieve sancties voor overtredingen in het verkeer. Zo werden overtredingen in verband met foutief stilstaan en parkeren onderdeel van de nieuwe GAS-wetgeving, en werden inbreuken op verkeersborden C3 (verboden toegang in beide richtingen) en F103 (begin voetgangerszone) toegevoegd. Het gaat hier telkens om inbreuken van de vierde soort. Deze inbreuken blijven ingeschreven in het Strafwetboek en kunnen zowel administratiefrechtelijk als strafrechtelijk worden bestraft (Schram & Lievens, 2015).

In dit onderzoek staan de GAS-boetes voor verkeersinbreuken inzake foutief stilstaan en parkeren centraal. Omwille van het specifieke karakter van parkeerovertredingen werden er bijzondere regels en procedures rond opgesteld in het Koninklijk Besluit van 9 maart 2014. Ten eerste, hanteren alle lokale besturen dezelfde vaste geldboetes, toegepast op inbreuken begaan door personen vanaf 18 jaar. Ten tweede, worden twee categorieën van overtredingen onderscheiden.

[11] In 2004 bijvoorbeeld, werden een reeks strafrechtelijke, kleine misdrijven gedepenaliseerd, werd de GAS-regeling uitgebreid naar jongeren tussen 16 en 18 jaar en kwam er een systeem van bemiddeling op lokaal niveau. In 2013 zag een nieuwe, verstrengde en op zichzelf staande GAS-wet het levenslicht die verdere aanpassingen doorvoerde zoals onder meer het verhogen van de boetes, het verlagen van de leeftijdsgrens naar 14 jaar, de uitbreiding van de lijst met inbreuken en het aantal vaststellende ambtenaren, en de vereenvoudiging van de procedure

- Categorie 1 overtredingen omvatten bijvoorbeeld het onjuist gebruik van de parkeerschijf en het overtreden van een verbod voor (stilstaan of) parkeren (meer specifiek het stilstaan en parkeren op woonerven, het parkeren voor de inrit van een eigendom, het parkeren in voetgangerszones, in de tegengestelde rijrichting, of op hinderlijke en gevaarlijke plaatsen). De boete bedraagt € 58.

- Categorie 2 heeft bijvoorbeeld betrekking op overtredingen rond het parkeren op het voetpad, zebrapad en fietspad, of op plaatsen voor personen met een handicap zonder de officiële parkeerkaart. De boete bedraagt € 116.

Alle andere parkeerinbreuken die niet meer strafbaar zijn volgens het strafwetboek worden niet via een GAS-boete bestraft en maken geen deel uit van dit onderzoek.[12]



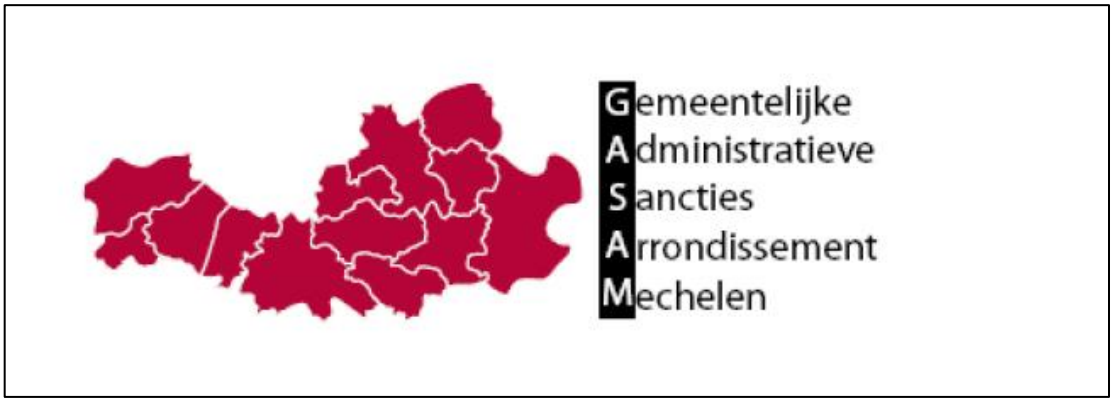
Figuur 14. Verkeersbord parkeerverbod (E1)

4.1.2 GASAM procedures

Om de algemene GAS-wetgeving voor overlast en verkeer in het arrondissement Mechelen toe te passen, werd in 2008 de interlokale vereniging GASAM opgericht. Het beheer van GASAM is toevertrouwd aan een beheerscomité.[13] Elke gemeente heeft één lid van het college van burgemeester en schepenen aangeduid als effectieve vertegenwoordiger en één lid als plaatsvervanger. De intergemeentelijke vereniging voor de ontwikkeling van Gewest Mechelen en Omgeving (IGEMO) maakt met adviserende stem deel uit van het beheerscomité (GASAM, 2020).

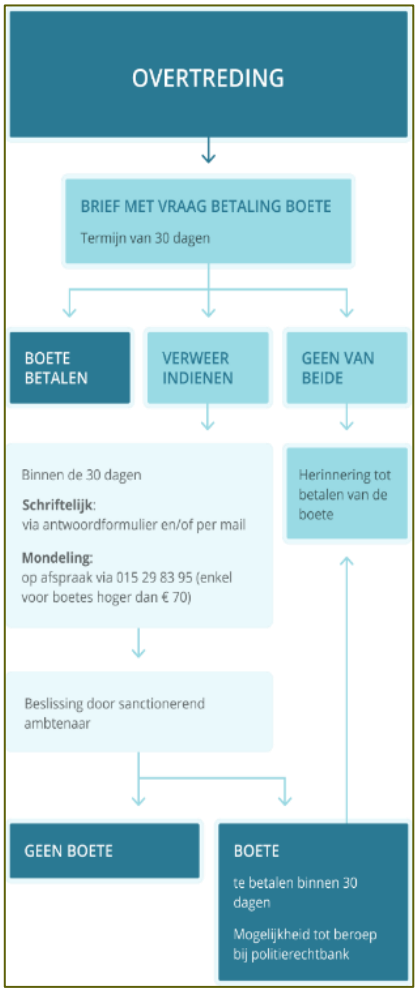
[12] In dat geval wordt een parkeerretributie opgelegd, zoals bijvoorbeeld bij het niet (voldoende) betalen in een zone betalend parkeren, het vergeten van de parkeerschijf in een blauwe zone, of de maximale parkeertijd overschrijden. Parkeerwachters stellen zulke gedepenaliseerde overtredingen vast en reiken de parkeerretributies hiervoor uit. De retributiebedragen liggen ook lager dan die voor een GAS-boete parkeren of stilstaan.
[13] Volgende gemeenten zijn opgenomen in GASAM en zijn vertegenwoordigd in het beheerscomité: Mechelen, Berlaar, Bonheiden, Bornem, Duffel, Heist-op-den-Berg, Lier, Nijlen, Puurs-Sint-Amands, Sint-Katelijne-Waver, Putte en Willebroek

De GASAM procedure is in detail beschreven in Figuur 16. In wat volgt bespreken we in grote lijnen de passages die relevant zijn het tijdig betalen van de boetes. Alles begint met de politie die een overtreding vaststelt en een proces-verbaal (PV) opmaakt en doorstuurt naar de sanctionerend ambtenaar van de Stad Mechelen, die vervolgens beoordeelt of de overtreding voldoende bewezen is en er een boete mag worden opgelegd. De sanctionerend ambtenaar stuurt binnen de 15 dagen na ontvangst van de PV een (niet-aangetekende) brief met daarin informatie over de vastgestelde feiten, de inbreuk en het bedrag van de administratieve sanctie naar de overtreder (GASAM, 2020). De overtreder dient de boete vervolgens binnen de 30 dagen te betalen, tenzij de overtreder binnen dezelfde termijn een verweer indient via een bij de boetebrief gevoegd antwoordformulier.[14] Indien de boete niet binnen de termijn wordt betaald dan volgen er in totaal drie herinneringen (één bij gewone zending, één bij aangetekende zending zonder extra kosten en één bij aangetekende zending met € 20 extra kosten). Nadien volgt invordering via de gerechtsdeurwaarder, waardoor de kosten voor de overtreder, inclusief BTW, kunnen oplopen tot € 197.



Figuur 15. Logo GASAM

[14] De overtreder kan binnen deze termijn op zijn verweer gehoord worden wanneer het bedrag van de administratieve geldboete hoger ligt dan 70 euro. Indien het verweer wordt aanvaard, ontvangt de overtreder een schriftelijke bevestiging en wordt het dossier afgesloten. Indien het verweer niet wordt aanvaard dan ontvangt de overtreder een schriftelijk bericht en moet de boete binnen een nieuwe termijn van 30 dagen betaald worden. Binnen de nieuwe termijn bestaat de mogelijkheid om in beroep te gaan bij de politierechtbank.

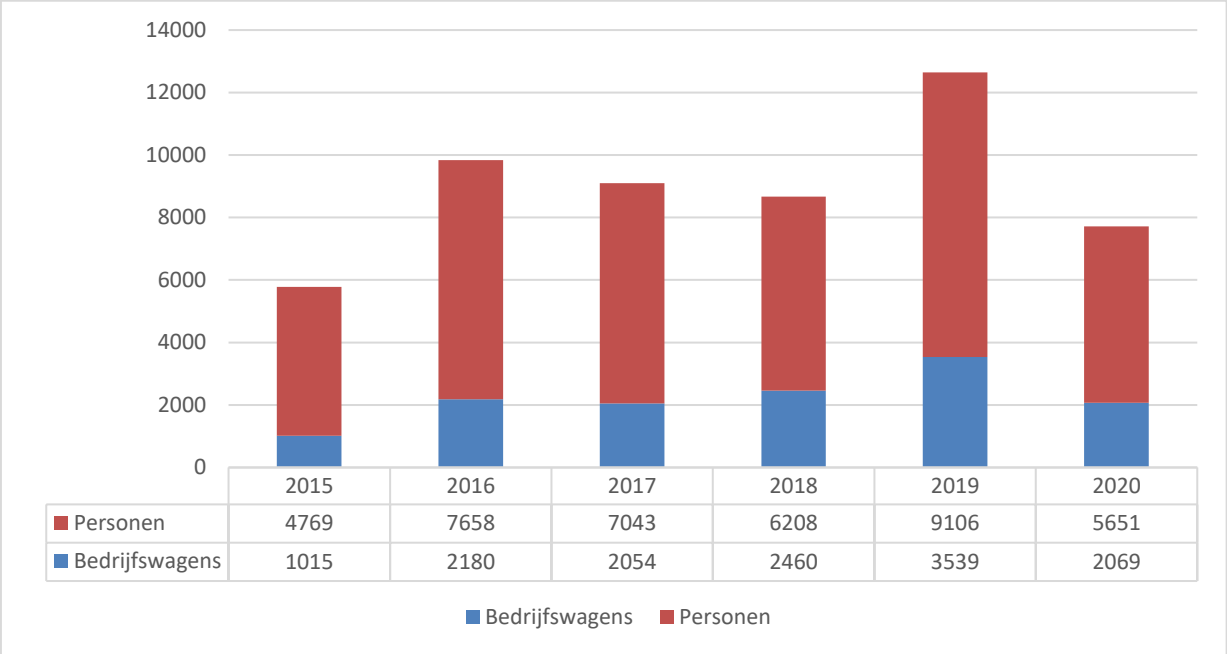


Figuur 16. Procedures GASAM

4.1.3 GASAM cijfers

Voor de periode 2015 tot 2019 stellen we in Figuur 17 vast dat er per jaar gemiddeld 9206 personen en organisaties een GAS-boete ontvangen van de Stad Mechelen voor fout parkeren of stilstaan (GASAM, 2021). Voor dit onderzoek nemen we 2019 als referentiejaar. Dat jaar was er een opvallende stijging merkbaar tot 12645 dossiers. De meest voorkomende inbreuk in 2019 betreft een overtreding waar een verkeersbord E1

(parkeerverbod, zie Figuur 14) is opgesteld, goed voor 46% van de overtredingen.[15] Nog in 2019 werden 72% van de boetes uitgedeeld aan individuele personen zonder bedrijfswagen en 28% aan personen met een bedrijfswagen (GASAM, 2020).[16]



	2015	2016	2017	2018	2019	2020
Personen	4769	7658	7043	6208	9106	5651
Bedrijfswagens	1015	2180	2054	2460	3539	2069

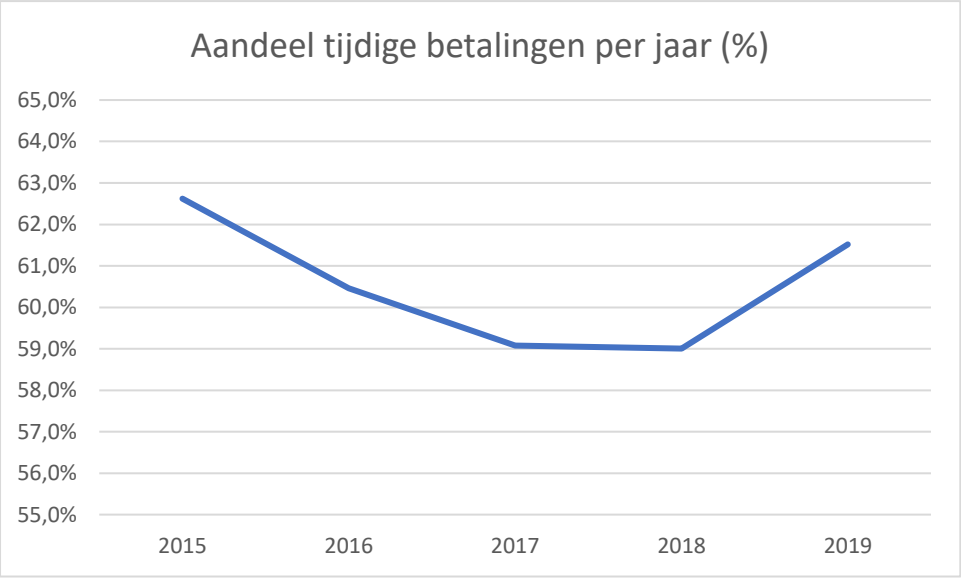
Figuur 17. Aantal GASAM dossiers parkeren en stilstaan per jaar (2015-2020)

Voor dit experiment laten we de bedrijfswagen buiten beschouwing en nemen we uitsluitend de boetes voor individuele personen mee in de analyse. We zijn immers voornamelijk geïnteresseerd in het individuele beslissingsproces van burgers en voor wagens van ondernemingen kunnen we het betalingsproces in mindere mate rechtstreeks relateren aan één individuele persoon. Als we enkel op personen selecteren en vervolgens de verhouding van het aantal tijdig betaalde boetes in kaart brengen, komen we tussen 2015 en 2019 uit op een gemiddelde tijdige betalingsgraad van 60,6% (zie Figuur 18). Dit betekent dat jaarlijks

[15] Het aantal dossiers dat in 2019 werd opgestart en waar geen verweer voor werd ingediend bedraagt 91%. Van het aantal dossiers waarvoor een verweer werd ingediend werd 34% gunstig beslist. In 2019 werd tegen negen dossiers een beroep ingesteld bij de politierechtbank

[16] In het datasysteem van GASAM wordt een onderscheid gemaakt tussen houders van een kentekenplaat die een rijksregister heeft (natuurlijk persoon) of een KBO nummer (rechtspersoon). Hier geven we dit onderscheid weer als het verschil tussen overtreders zonder en met bedrijfswagens. Al moeten we opmerken dat uitzonderingen niet uitgesloten zijn. Zo kunnen burgers wagens met een KBO nummer als natuurlijke persoon leasen.

ongeveer 30% tot 35% van de overtreders hun boete te laat betalen (GASAM, 2019). Het is die betalingsgraad die we in dit onderzoek door middel van nudging in de boetebrief willen laten stijgen.



Figuur 18. Percentages tijdige betalingen per jaar (2015-2019)

4.2 NUDGING EN BELASTINGEN

De aandacht voor informatie en communicatie in het beleidsinstrumentarium neemt de laatste jaren sterk toe. In die mate zelfs dat John (2017) zonder veel tegenkanting kan stellen dat vandaag alle tools van de overheid "informatiegebonden" zijn. We hypothetiseren dat door boetebrieven te wijzigen op basis van "communicatieve nudges" – i.e. subtiele, gedragsgeïnspireerde aanpassingen in de informatie-omgeving - we ook het aantal tijdige betalingen kunnen stimuleren.

Voor de term "communicatieve nudges" bouwen we verder op de inzichten van Esmark (2019, p. 142, eigen vertaling) die stelt dat "gedrag kan gewijzigd, geleid en aangemoedigd kan worden door informatie en communicatie centraal te plaatsen in het ontwerp van de keuze-architectuur rond *welfare choices*". We vonden echter weinig voorbeelden van nudging in de context van boetes. Een studie die we meenemen in ons onderzoek is het Britse RCT veldexperiment van Haynes et al. (2013) rond zogenaamde "*delinquent*

fines", boetes die door een rechtbank zijn opgelegd. Dat onderzoek heeft aangetoond dat het sturen van herinneringsberichten per sms op zichzelf al effectief kan zijn voor het betalen van de boete, en dat het personaliseren van het bericht dit effect nog kon versterken. Een aantal exploratieve studies uit Zuid-Afrika tonen verder dat het gebruik van sociale normen en het duidelijk communiceren over de gevolgen bij niet betalen van verkeersboetes een positief effect zouden kunnen hebben op het betalingsgedrag (Bantjes et al., 2021; du Plessis et al., 2020).

Omwille van de beperkte wetenschappelijke aandacht en empirische bewijskracht voor het nudgen van boetebrieven inspireren we ons voor deze studie in grote mate op een ander specifiek beleidsdomein, dat van de belastingen. Het wijzigen van de keuze-architectuur via het 'tweaken' van belastingbrieven is populair aangezien dit soort aanpassingen relatief haalbaar, betaalbaar en opschaalbaar zijn. Sterk uitgebouwde en gemoderniseerde belastingdiensten bieden ook troeven op het vlak van randomisatie en replicatie (John & Blume, 2018). Daarnaast spreken we op het fiscale terrein al snel over grote steekproeven en hoge betalingsvolumes, waardoor zelfs kleine effecten in absolute cijfers een grote (financiële) impact kunnen hebben.

4.2.1 Afschrikken, herschikken of op de moraliteit mikken?

Pogingen om het tijdig en correct betalen van belastingen (*tax compliance*) te stimuleren, zijn onlosmakelijk verbonden met het innen van belastingen. De voorbije decennia explodeerde het theoretisch en empirisch onderzoek naar het gedrag van de belastingbetaler (Slemrod & Weber, 2012). Het academische debat over dit thema richtte zich in de beginjaren vooral op belastingontduiking (*tax evasion*) (bijvoorbeeld: Srinivasan, 1973). Allingham & Sandmo (1972) vertrokken vanuit een rationeel economisch gedragsperspectief om te verklaren waarom mensen belastingen betalen of ontduiken. Het zogenaamde Allingham-Sandmo (AS) model voorspelt dat belastingontduiking vermindert wanneer de kans om ontdekt en bestraft te worden toeneemt, en/of wanneer de extra kosten en boetes toenemen. Ook de hoogte van de belastingen in verhouding tot het inkomen speelt een rol. Het model steunt op de afweging tussen kosten en baten, en beargumenteert dat burgers voor belastingontduiking kiezen omdat de kosten van belastingen betalen hoger worden ingeschat dan de kosten om geen belastingen te betalen (Jensen et al., 2018; Milliron & Toy, 1988).

De voorbije twee decennia mengen ook onderzoekers met een gedragseconomische insteek zich in het debat en wijzen op het belang van psychologische factoren, zoals de gepercipieerde morele kosten, de betalingsgewoonten en de houding van het individu ten opzichte van belastingen in het algemeen (Kirchler, 2007). Daarnaast wordt in het wetenschappelijk onderzoek steeds meer aandacht geschonken aan de omstandigheden waarbinnen de belastingplichtige een keuze maakt. Onder meer de beschikbare informatie, de complexiteit van het betalingsproces en het gebruik van reminders werden onderzocht (Hallsworth, 2014; Jensen et al., 2018; Torgler, 2007). In het kielzog van de gedragsbenadering ging de aandacht ook uit naar de toepassing van nieuwe theorieën in het onderzoek naar belastingen, zoals de prospect theorie (Kahneman & Tversky, 1979), cognitieve biasen zoals verliesaversie (*loss aversion*) en een reeks morele boodschappen, zoals het gebruik van descriptieve, sociale normen (Botzem, 2019). Via deze weg slopen nudges binnen in de werking van de belastingdiensten.

Antinyan & Asatryan (2019) onderscheiden twee categorieën nudge-boodschappen op basis van een literatuuronderzoek en meta-analyse naar 40 studies over nudging in een belastingcontext. De eerste categorie nudge-boodschappen volgen een *deterrence* benadering en gebruikt afschrikwekkende nudges. De tweede categorie volgt een *non-deterrence* benadering en steunt op niet-afschrikwekkende nudges. Deze onderverdeling resoneert met de twee stromingen, de rationele en de gedragskundige, die we hierboven al schetsten (Hallsworth, 2014).

- De *deterrence* benadering beschouwt belastingplichtigen als rationele en nutsmaximaliserende burgers die kosten en baten afwegen. Een afschrikwekkende nudge vanuit de belastingdienst communiceert over het niet correct betalen van belastingen en legt daarbij de nadruk op de pakkans en mogelijke extra kosten. Het louter verhogen van deze kosten zelf is geen nudge, maar een financieel beleidsinstrument. De nudge uit zich via de manier waarop deze kosten gecommuniceerd, geaccentueerd en geframed worden (Jensen et al., 2018). Volgens dezelfde redenering is het verhogen van het aantal audits en controles geen nudge, terwijl het communiceren en vooral het accenturen van de pakkans bij verhoogde audits en controles wel als een nudge kan beschouwd worden (Alm et al., 2009; Chirico et al., 2019; Slemrod et al., 2001).

- De non-deterrence benadering kijkt naar andere factoren dan economische en vertrekt vanuit de assumptie dat belastingplichtigen uit zichzelf geneigd zijn om belastgingen te betalen. De vraag die de burger zich stelt, is dan "Wat moet ik doen om mijn belastingen correct te betalen?" en niet "Wat moet ik doen om ongestraft belastingen te ontwijken?" (Mcgraw & Scholz, 1991). Nudges worden dan toegepast om belastingbetalers zo gemakkelijk mogelijk hun belastingen tijdig en correct te laten betalen. Verder kunnen nudges ook morele argumenten aanhalen en een beroep doen op burgerplicht (Kirchler, 2007). Deze niet-afschrikwekkende nudges maken dus geen gebruik van dreiging. Voorbeelden zijn morele boodschappen, berichten die inspelen op *public fairness*, *public goods* booschappen die het belang van belastingen voor publieke dienstverlening benadrukken en sociale normen die zowel descriptief als normatief kunnen zijn (Antinyan & Asatryan, 2019; De Neve et al., 2021; John & Blume, 2018; Vainre et al., 2020).

Om de effectiviteit van deze twee soorten boodschappen en nudges te testen en empirisch te onderbouwen wordt op methodologisch vlak steeds vaker gebruik gemaakt van veldexperimenten op basis van RCT designs (Hallsworth, 2014).

4.2.2 Britse veldexperimenten

De eerste en allicht bekendste toepassing van zulke veldexperimenten met nudge-interventies vond plaats binnen de Britse belastingdienst, *Her Majesty's Revenue and Customs* (HMRC). Die zette in 2011 in samenwerking met het Britse *Behavioural Insights Team* (BIT) een grootschalig RCT veldexperiment op met meer dan 100.00 burgers. Deze burgers kregen een aanmaningsbrief omdat ze te laat waren met het terugbetalen van hun belastingen (Hallsworth et al., 2017). Het loont de moeite om dit Britse veldexperiment inzake het tijdig terug betalen van belastingschulden nader te bestuderen met het oog op dit onderzoek. Het veldexperiment maakte gebruik van nudge-boodschappen met een focus op sociale normen enerzijds en op het belang van publieke dienstverlening anderzijds. Die werden gerandomiseerd verstuurd op basis van unieke dossiernummers naar zes groepen belastingplichtigen. Van de zes groepen ontving één controlegroep de 'oude' brief. De vijf andere groepen ondergingen de experimentele manipulatie. Er werd meer bepaald een korte zin toegevoegd vlak na de openingszin die voor elke groep anders was. De vijf interventiezinnen en hun relatieve effect of de tijdige betalingsgraad zijn weergegeven in Tabel 11.

Tabel 11. Interventiezinnen en resultaten eerste veldexperiment VK (Hallsworth et al., 2017)

Groep	Zin	Effect
Controle	/	
Sociale norm basis	Negen op tien mensen betaalt zijn belastingen op tijd.	+1,3%***
Sociale norm land	Negen op tien mensen in het VK betaalt zijn belastingen op tijd.	+2,1%***
Sociale norm minderheid	Negen op tien mensen in het VK betaalt zijn belastingen op tijd. U behoort momenteel tot een erg kleine minderheid die ons nog niet heeft terugbetaald.	+5,1%***
Winstframe publieke dienstverlening	Het betalen van belastingen zorgt ervoor dat we allemaal van publieke diensten kunnen genieten zoals de NHS, openbare wegen en scholen.	+1,6%***
Verliesframe publieke dienstverlening	Het niet betalen van belastingen zorgt ervoor dat we een aantal publieke diensten kunnen verliezen, zoals de NHS, openbare wegen en scholen.	+1,6%***

Alle aanmaningsbrieven hadden een significant effect op het sneller betalen van achterstallige belastingen, zowel na 23 dagen (laatste dag voor eerste reminder) als na 48 dagen (moment dat een callcenter wordt ingeschakeld voor achterstallige schulden). De vijf boodschappen samen hadden een gemiddeld effect van 2.2 % in vergelijking met de controlegroep ($p<0.001$). De sociale norm boodschap met een focus op "behoren tot de minderheid" had daarbij het meeste effect. Ook de framing rond publieke dienstverlening versnelde de terugbetalingen, maar hier was geen significant verschil tussen een winstframe of een verliesframe. Op basis van alle vijf de boodschappen werd naar schatting £4,9 miljoen sneller terugbetaald. Het gemiddelde effect van 2.2% lijkt dus op het eerste zich beperkt, maar kan aanzienlijk zijn in absolute bedragen.

Een jaar nadien werd een nieuw veldexperiment met dezelfde basiskenmerken uitgevoerd bij ongeveer 120.00 inwoners in Engeland, Wales en Noord-Ierland. Hierbij werd de sociale norm met een focus op het "behoren tot een minderheid" gerepliceerd, werden descriptieve en injunctieve sociale normen met elkaar vergeleken, en werd de impact van het toevoegen van financiële informatie getest. Het experiment bestond deze keer uit 13 groepen en één controlegroep. De interventies en resultaten van dit experiment staan weergegeven in Tabel 12.

Tabel 12. Interventiezinnen en resultaten tweede veldexperiment VK (Hallsworth et al., 2017)

Groep	Zin	Effect
Controle	/	
Descriptieve norm basis	De grote meerderheid van de mensen in het VK betaalt zijn belastingen op tijd.	+1,4%*
Descriptieve norm lokaal	De grote meerderheid van de mensen in jouw buurt betaalt zijn belastingen op tijd.	+2,2%***
Descriptieve norm schulden	De meeste mensen die evenveel moeten terugbetalen als u, hebben ondertussen betaald.	+3.0%***
Descriptieve norm schulden + buurt	De grote meerderheid van de mensen in jouw buurt betaalt zijn belastingen op tijd. De meeste mensen die evenveel moeten terugbetalen als u, hebben ondertussen betaald.	+5.0%***
Minderheid status	U behoort momenteel tot een erg kleine minderheid die ons nog niet heeft terugbetaald.	+4.7%***
Descriptieve norm minderheid	Negen op tien mensen in het VK betaalt zijn belastingen op tijd. U behoort momenteel tot een erg kleine minderheid die ons nog niet heeft terugbetaald.	+4.2%***
Morele plicht	Iedereen in het VK zou zijn belastingen op tijd moeten betalen.	+2.2%***
Algemene, injunctieve norm	De grote meerderheid van de mensen is akkoord dat iedereen in het VK zijn belastingen op tijd zou moeten betalen.	+0.6%
Fractie injunctieve norm	Negen op tien mensen is akkoord dat iedereen in het VK zijn belastingen op tijd zou moeten betalen.	+1.7%**
Percentage injunctieve norm	88% van de mensen is akkoord dat iedereen in het VK zijn belastingen op tijd zou moeten betalen.	+3.4%***
Injunctieve + descriptieve norm	Negen op tien mensen is akkoord dat iedereen in het VK zijn belastingen op tijd zou moeten betalen. En negen op tien mensen betalen ook op tijd.	+3.6%***
Extra info	U kan betalen via debet kaart, credit kaart of Direct Debit. Je kan ook rechtstreekst via het internet en uw telefoon banking werken. Voor meer informatie, zie …	+3.2%***
Interesten	Wij zullen u interesten op dit bedrag aanrekenen.	+3.9%***

De 13 boodschappen samen hadden een gemiddeld effect van 2.9% na 23 dagen in vergelijking met de controlegroep ($p<0.001$). De resultaten rond descriptieve sociale normen werden van het eerste experiment werden bevestigd en de combinatie van schulden en buurtgevoel scoorde het hoogste van alle boodschappen (+5%). Zowel injunctieve als descriptieve normen hebben een significant effect, maar het effect van de descriptieve normen was groter. Ook het leveren van extra informatie over de betalingswijze en de expliciete waarschuwing voor mogelijke interesten deed de betalingssnelheid toenemen. Voor dit tweede experiment werd berekend dat de dertien boodschappen instonden voor £9,3 miljoen aan snellere terugbetalingen (Hallsworth et al., 2017). Geen van beide experimenten had echter betrekking op nieuwe inkomsten. De onderzoeken draaiden rond de versnelde betalingen van reeds geboekte belastinginkomsten. Een groot deel van deze schulden zouden op iets langere termijn waarschijnlijk ook in de staatskas beland zijn. Het sneller en tijdig innen van belastingen heeft evenwel enkele voordelen. Overheden hoeven minder administratieve handelingen te verwerken om de inning van schulden te regelen en kunnen daarmee kosten besparen. Burgers lopen daarnaast minder kans op hogere kosten en ervaren minder financiële stress (Hallsworth et al., 2017).

4.2.3 Belgische veldexperimenten

In België liet de FOD Financiën zich door het Britse voorbeeld inspireren om vanaf 2015-2016 samen met gerenommeerde gedragseconomen grootschalige, wetenschappelijk onderbouwde RCT-veldexperimenten op te zetten. Het onderwerp van deze eerste Belgische veldexperimenten betrof de aanmaningsbrief voor de personenbelasting. Jaarlijks moeten ongeveer 1,8 miljoen burgers bijbetalen, goed voor een bedrag van 3,6 miljoen euro (Luts & Van Roy, 2018). Burgers vernemen dit via hun aanslagbiljet en krijgen vervolgens twee maanden de tijd om het openstaande saldo te betalen. Jaarlijks zijn er echter meer dan 250.000 laattijdige betalers. De steekproef in de Belgische studie bestond uit 225.877 burgers. Het design van het onderzoek werd opgebouwd rond negen gerandomiseerde groepen die elk een andere boodschap te lezen kregen. De randomisatie gebeurde op basis van de laatste twee cijfers in het nationale nummer van de belastingplichtige. Een contolegroep ontving de historisch gebruikte herinneringsbrief (de "Ancien" genoemd). Deze bevatte financiële informatie (o.a. het openstaande saldo) en administratief taalgebruik om de ontvanger te wijzen op zijn plichten en mogelijke bijkomende kosten. Deze brief werd vereenvoudigd en als basisbrief gebruikt voor de acht interventiegroepen. De nieuwe basisbrief bevatte vijf aanpassingen: eenvoudiger taalgebruik, persoonlijke aanspreking, gebruik van kleur, actiegerichte kaders en positieve

bekrachtiging via een bedanking voor (eventueel) reeds gedane betaling. Hij werd in deze brief naar de eerste interventiegroep verstuurd. De overige interventiegroepen ontvingen de nieuwe basisbrief met nudge boodschappen eraan toegevoegd. In Tabel 13 hebben we de interventies en effecten van dit experiment weergegeven.

Tabel 13. Interventiezinnen en resultaten veldexperiment België (De Neve et al., 2021; Luts & Van Roy, 2018)

Groep/Type brief	Interventie (zin)	Effect
Controle	/	
Vereenvoudiging	Taal, persoonlijk, kleur, kader en bedanking	+9.1%
Vereenvoudiging + Sociale norm	In België worden 95% van de belastingen op tijd betaald.	+9.2%
Vereenvoudiging + Public goods negatief frame	Zonder belastinginkomsten kunnen publieke diensten zoals gezondheid, onderwijs en ordehandhaving niet langer gegarandeerd worden.	+8.8.%
Vereenvoudiging + Public goods positief frame	Belastinginkomsten zorgen ervoor dat basis publieke diensten zoals gezondheid, onderwijs en ordehandhaving functioneren.	+7.8%
Vereenvoudiging + Sociale norm + Public Goods	Een combinatie van de twee bovenstaande boodschappen.	+8.4%
Vereenvoudiging + Explicit penalty	De kosten voor te laat te betalen bedragen gemiddeld 209 euro en kunnen, afhankelijk van de omstandigheden, nog verder oplopen.	+10.9%
Vereenvoudiging + Actieve keuze	Het niet betalen van belastingen zal beschouwd worden als een bewuste keuze.	+9.3%
Vereenvoudiging + Actieve keuze + Explicit penalty	Een combinatie van de twee bovenstaande boodschappen.	+10.6%

Alle vereenvoudigde brieven geven een positief significant effect op de terugbetalingsgraad na 14 dagen ten opzichte van de controlebrief. Het vereenvoudigen van de brieven heeft een duidelijke impact. Het meest effectief ten opzichte van de 'gewone' vereenvoudigde brief zijn de afschrikkingsboodschappen (*deterrence*) zoals in de Explicit Penalty brief (al dan niet gecombineerd met Actieve keuze). De public goods en sociale norm boodschappen hebben dan weer geen bijkomend effect ten opzichte van de vereenvoudigde brief, of zelfs een negatief effect in vergelijking met de nieuwe basisbrief (niet in vergelijking met de controlebrief).

Het valt op dat de resultaten van de onderzoeken in België en het Verenigd Koninkrijk verschillen. Waar de Britten zich vooral lieten overtuigen door sociale norm en in mindere mate door public goods boodschappen, lijken Belgen vooral gevoelig voor het vermijden van extra kosten. Verder stellen we vast dat het vereenvoudigen van informatie zelf óók effectief is – met name doordat beoogde acties duidelijk worden en irrelevante en complexe randinformatie geëlimineerd.

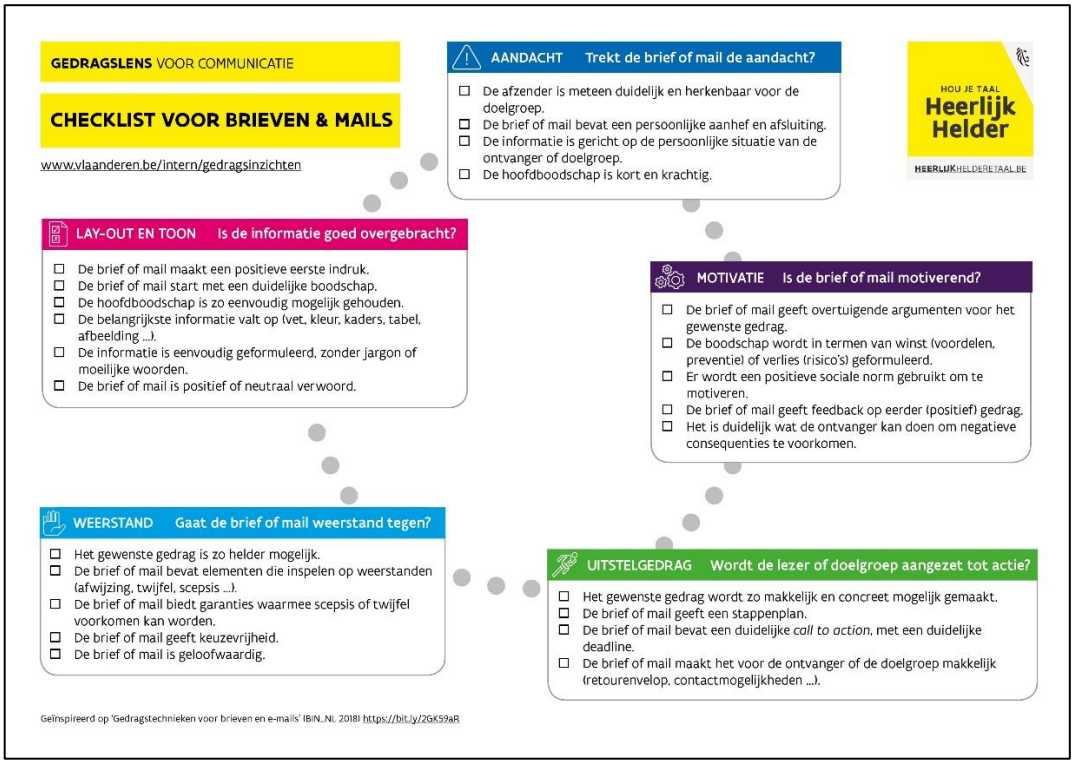
Daarnaast betrof het een kostenefficiënt project (Luts & Van Roy, 2018). De kostenposten personeel en materiaal bedroegen samen bijna € 80.000, terwijl de versnelde meeropbrengst na 14 dagen bijna € 22 miljoen bedroeg. Omdat een groot deel van dit bedrag op langere termijn waarschijnlijk ook zou zijn terugbetaald, gaat het hier niet om extra inkomsten, maar om snellere inkomsten. Het voordeel ligt vooral in het verminderen van de interne kosten en werklast bij de overheid, en het verminderen van mogelijke nalatigheidskosten aan de kant van de burger. De FOD Financiën optimaliseerde na 2016 de aanmaningsbrief verder via replicatiestudies en het finetunen van de boodschappen in de brieven. Zo werd het effect nog groter door een verliesframe te gebruiken voor de Explicit Penalty boodschap. Een jaar later werden zelfs bijkomende effecten geïdentificeerd die mogelijk ook op langere termijn kunnen blijven doorwerken. Belastingplichtigen die voor het aanslagjaar 2015 een aangepaste aanmaningsbrief hadden ontvangen, hadden bijvoorbeeld een kleinere kans om voor het aanslagjaar 2016 opnieuw laattijdig te betalen (Luts & Van Roy, 2018). De FOD Financiën verkende nog andere pistes, zoals aanpassingen aan het aanslagbiljet of het toevoegen van een taartgrafiek bij het elektronisch invullen van de belastingaangifte om visueel weer te geven waaraan de overheid het geïnde belastinggeld spendeert (De Neve et al., 2021).

4.3 NUDGING EN PARKEERBOETES: INTERVENTIES

De belastingliteratuur biedt ons handvaten om een gedragsanalyse uit te voeren en interventies te ontwerpen. Daarnaast hebben we ook een beroep gedaan op checklist voor brieven en mails van de Vlaamse Overheid[17] en een gelijkaardige handleiding met gedragstechnieken vanwege het Nederlandse *Behavioural Insights Newterk* (BIN-NL, 2018). We hebben de volgende vijf gedragsfactoren samen met het DKBZ Team Gedragsinzichten en met de medewerkers van GASAM stelden we onszelf de volgende vijf vragen,

[17] https://overheid.vlaanderen.be/gedragsinzichten-heldere-brieven-e-mails-en-webplatformen

overgenomen uit Figuur 19: Trekt de brief de aandacht? Is de brief motiverend? Wordt de lezer aangezet tot actie? Gaat de brief weerstand tegen? Is de informatie goed overgebracht?

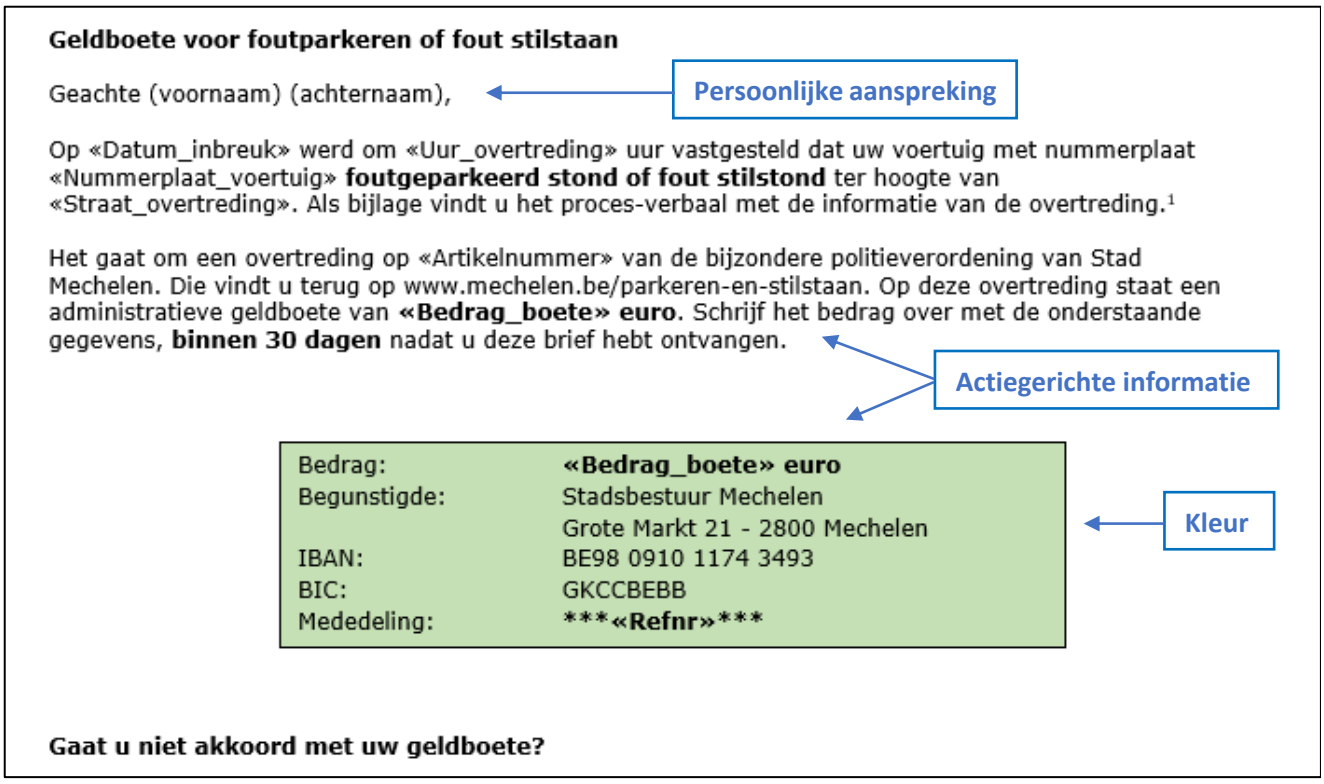


Figuur 19. Checklist voor brieven en mails Vlaamse Overheid

We kwamen tot de conclusie dat deze vijfgedragsfactoren – aandacht, motivatie, uitstelgedrag, weerstand en lay-out – allemaal belangrijke determinanten waren om van probleemgedrag naar doelgedrag te bewegen. Bovenop de voorbeelden van de Britse en Belgische veldexperimenten en het gebruik van nudges in een lokale belastingcontext (John & Blume, 2018; Larkin et al., 2019), bood de checklist een extra hulpmiddel om interventies te ontwerpen. Tot slot hielden we rekening met de vier principes van het EAST framework, ontworpen door BIT: om het gewenste gedrag te beïnvloeden hebben we het gedrag zo gemakkelijk (easy), aantrekkelijk/opvallend (attractive), sociaal (social) en tijdsgebonden (timely) en dus zo EAST mogelijk gemaakt (BIT, 2014). Zo kwamen we uit bij de volgende drie interventies.

4.3.1 Interventie 1: Maak de boodschap gemakkelijk en vereenvoudig de brief

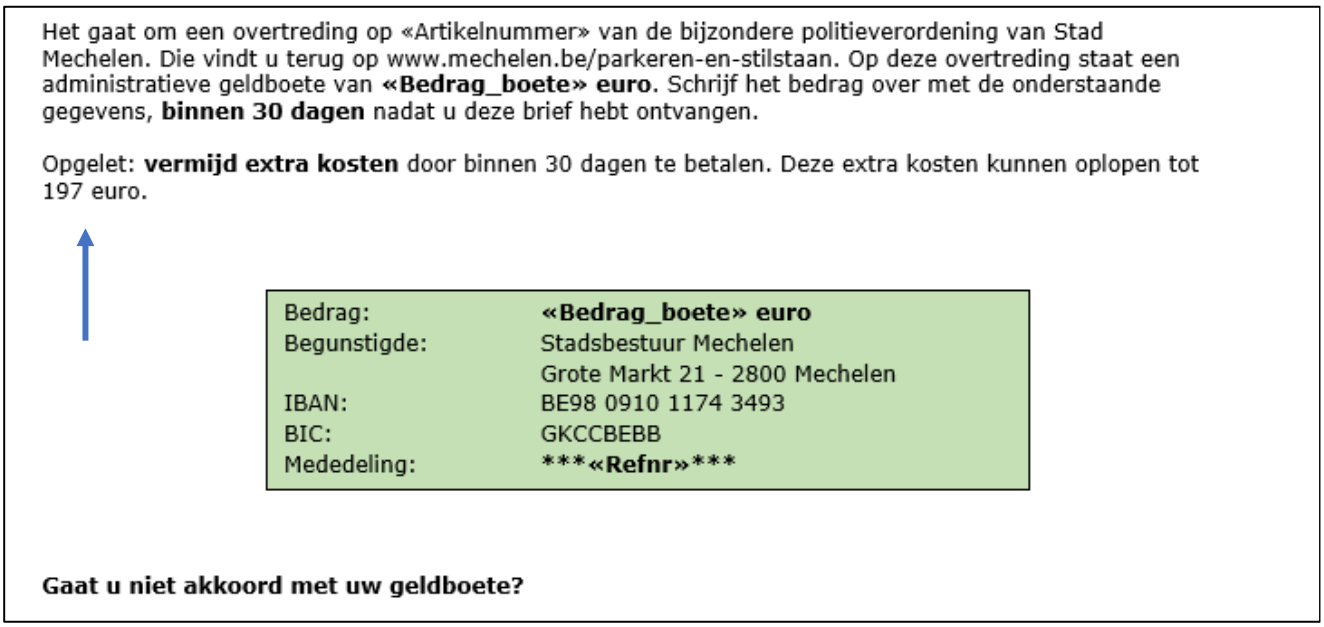
De eerste interventie die we hebben toegepast op de parkeerboetebrief is het vereenvoudigen van de boodschap. Verschillende studies rond belastingen hebben aangetoond dat de vereenvoudiging van brieven mensen kan helpen om sneller te betalen (De Neve et al., 2021). Vanuit die invalshoek hebben we de tekst laten screenen door het team gedragsinzichten (DKBZ), het team Heerlijk Helder van de Vlaamse overheid, en de dienst Marketing & Communicatie van de Stad Mechelen. Bedoeling was om de parkeerboetebrief korter en helderder te maken. Verder hebben we een persoonlijke aanspreking toegevoegd en gebruik gemaakt van een groene kleur om de actierelevante informatie te accenturen. Door de complexiteit van de boodschap te verminderen en de actierelevante informatie uit te lichten, wilden we in de eerste plaats het (cognitieve) gemak verhogen om het gewenste gedrag te stellen.



Figuur 20. Maak de boodschap gemakkelijk en vereenvoudig de brief

4.3.2 Interventie 2: Maak de boodschap opvallend en accentueer mogelijke extra kosten

Voor de tweede interventie behielden we de vereenvoudigde basisbrief en voegden we er een afschrikwekkende boodschap aan toe. Studies rond deze *deterrence* boodschappen hebben in uiteenlopende belastingcontexten de effectiviteit aangetoond van het belichten van mogelijke vervolgacties in geval van het aanhoudend niet (tijdig) betalen van het schuldsaldo (Brockmeyer et al., 2019; Cranor et al., 2020; De Neve et al., 2021; Hallsworth et al., 2017; Pomeranz & Vila-Belda, 2019; Slemrod, 2018). Op basis van de meta-analyse van Antinyan & Asatryan (2019) weten we dat afschrikwekkende boodschappen 1,5 tot 2,5 procentpunten effectiever zijn dan niet-afschrikwekkende boodschappen in het stimuleren van (tijdige) betaling. Dit werd ook bevestigd in het veldexperiment van de FOD Financiën, waar de Explicit Penalty boodschap de grootste impact had op de tijdige betaling van openstaande belastingen (Luts & Van Roy, 2018). Voor het veldexperiment hebben we GASAM de maximale, reële kosten laten berekenen indien een persoon de boete niet tijdig betaalt. De som van extra administratiekosten, deurwaarderskosten en BTW kan oplopen tot een bedrag van € 197. Naast deze informatie over mogelijke, extra kosten voegen we ook de formulering "vermijd extra kosten" toe om in te spelen op verliesaversie (De Neve et al., 2021).


Het gaat om een overtreding op «Artikelnummer» van de bijzondere politieverordening van Stad Mechelen. Die vindt u terug op www.mechelen.be/parkeren-en-stilstaan. Op deze overtreding staat een administratieve geldboete van **«Bedrag_boete» euro**. Schrijf het bedrag over met de onderstaande gegevens, **binnen 30 dagen** nadat u deze brief hebt ontvangen.

Opgelet: **vermijd extra kosten** door binnen 30 dagen te betalen. Deze extra kosten kunnen oplopen tot 197 euro.

Bedrag:	**«Bedrag_boete» euro**
Begunstigde:	Stadsbestuur Mechelen
	Grote Markt 21 - 2800 Mechelen
IBAN:	BE98 0910 1174 3493
BIC:	GKCCBEBB
Mededeling:	***«Refnr»***

Gaat u niet akkoord met uw geldboete?

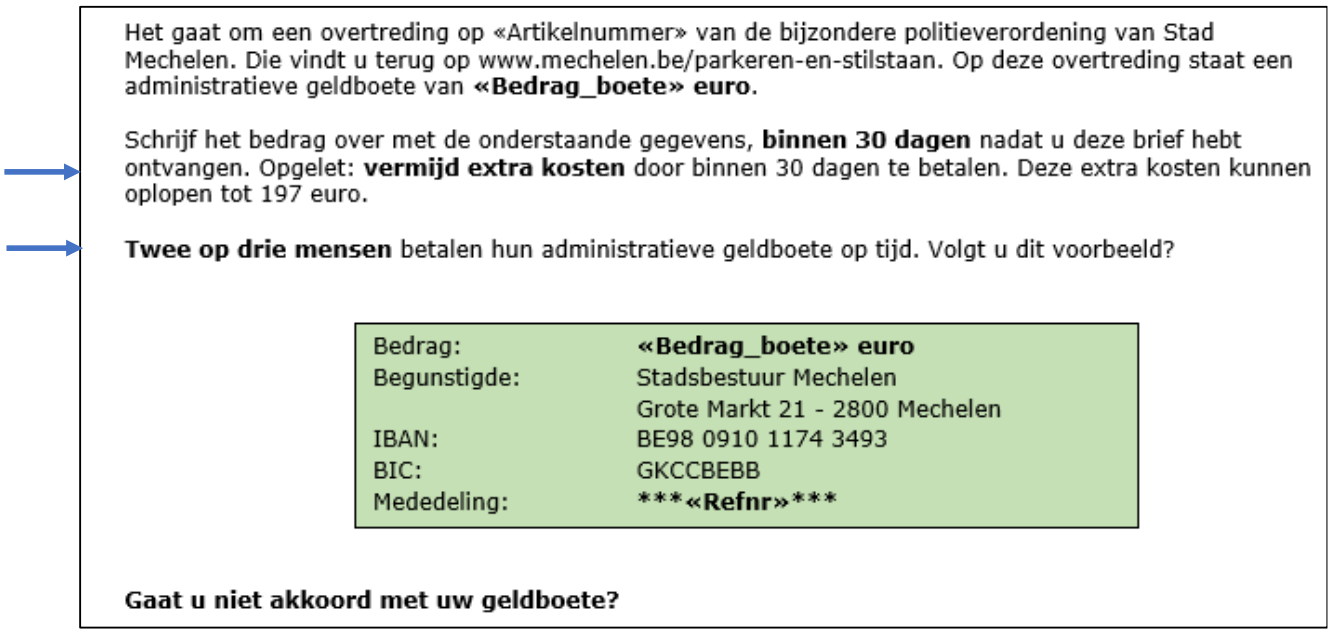
Figuur 21. Maak de boodschap opvallend en accentueer mogelijke extra kosten

4.3.3 Interventie 3: Maak de boodschap sociaal en beschrijf wat de meeste mensen doen

Voor de derde interventie selecteerden we een niet-afschrikwekkende (non-deterrence) boodschap op basis van een descriptieve sociale norm. We selecteerden geen morele boodschappen die inspelen op *public fairness* en *public goods* omdat die meer van toepassing zijn in een belastingcontext dan in een boetecontext. De descriptieve sociale norm wijst op de keuzes die de meerderheid van de personen in eenzelfde situatie maakt, en stimuleert mensen om die keuze te volgen. Sociale normen hebben hun nut getoond in verschillende beleidsdomeinen zoals afval sorteren (Schultz, 1999), zwerfvuil (Cialdini, 2003), energieverbruik (Allcott, 2011), waterconsumptie (Ferraro & Price, 2013) en het voorschrijven van antibiotica door artsen (Hallsworth et al., 2016). In het domein van belastingen vinden we echter gemengde resultaten. We zagen al dat bij het veldexperiment in het VK de sociale normen het meeste potentieel hadden (Hallsworth et al., 2017), en bij trials rond lokale belastingen zijn er ook positieve effecten vastgesteld (Larkin et al., 2019). Andere studies resulteerden dan weer in niet-significante effecten (Cranor et al., 2020; De Neve et al., 2021) of zelfs onverwachte negatieve effecten (Ariel, 2012; John & Blume, 2018).

Onze preliminaire compliance analyses op basis van de ruwe GASAM-data voor het jaar 2019 wezen op 65% tijdige betalingen. Hoewel deze sociale norm niet uitgesproken hoog ligt, hebben soortgelijke studies aangetoond dat deze effectief kan zijn. Zo blijkt bijvoorbeeld uit een grootschalig veldexperiment van de belastingdienst in Guatemala i.s.m. de Wereldbank dat het gebruik van een sociale norm van 64,5% een significant effect kan hebben op de betalingsgraad (Kettle et al., 2016).

In onze studie pasten wij daarom de volgende sociale norm toe: "*Twee op de drie mensen betalen hun administratieve geldboete op tijd. Volgt u dit voorbeeld?*"



Figuur 22. Maak de boodschap sociaal en beschrijf wat de meeste mensen doen

4.4 EEN GECLUSTERD DESIGN

In deze studie onderzochten wij of verschillende soorten communicatieve nudges effectief kunnen zijn om het aantal burgers dat tijdig betaalt te verhogen. Onze afhankelijke variabele (AV) was het aantal personen dat de geldboete voor fout parkeren of stilstaan op tijd heeft betaald. De verschillende soorten nudgeboodschappen vormden de interventies van ons experiment. Door de GASAM procedures konden we voor dit experiment de interventies niet 'toedienen' op basis van individuele randomisatie, maar waren we genoodzaakt om te randomiseren op basis van clusters waarbinnen individuele eenheden gegroepeerd zijn. In dit luiken lichten we de details van de afhankelijke variabele, de interventies en het specifieke design verder toe.

4.4.1 Het bepalen van de betalingsgraad

Eerst moeten we bepalen wat "op tijd betaald" betekent. De officiële betalingstermijn is 30 dagen. De officieuze betalingstermijn in het monitoringsysteem van GASAM bedraagt 32 dagen na het versturen van de boetebrief. Vanaf de 33ste dag wordt een betaling als "te laat betaald" gecategoriseerd. Zowel boetes die

te laat betaald worden als boetes die na 32 dagen nog openstaan vallen onder "te laat betaald". Om de evolutie in het aantal tijdig betaalde boetes te identificeren, meten we de verhouding tussen dossiers die "op tijd betaald" zijn en dossiers die "te laat betaald" werden.[18] Deze ratio noemen we de *betalingsgraad*.

We testten de effectiviteit van de verschillende interventies via vier groepen. De bestaande boetebrief (brief C) werd gebruikt als de controleconditie. De eerste interventie (brief V) betrof de vereenvoudiging van die controlebrief op basis de volgende aanpassingen: minder tekst, toegankelijke taal, persoonlijke aanspreking en kleur (zie uitleg hierboven). De tweede interventie (brief VA) nam de vereenvoudigde brief als basis en voegde daar een afschrikkende boodschap aan toe door te wijzen op mogelijke extra kosten indien het boetebedrag niet op tijd betaald zou worden. De derde interventie (brief VAS) combineerde de vereenvoudiging en de afschrikkende boodschap met een beschrijvende sociale norm, met name dat twee op drie personen de boete op tijd betaalt.

Het veldexperiment had een doorlooptijd van negen maanden, en liep van 1 juni 2020 tot en met 2 maart 2021. Op basis van het hoge aantal dossiers uit het verleden (n = 12.645 in 2019) was er oorspronkelijk een doorlooptijd van zes maanden voorzien voor het onderzoek. Het project werd echter verlengd met drie maanden door het achterblijven van het aantal vastgestelde GAS-boetes in 2020 (waarschijnlijk door de gevolgen van de COVID-19 maatregelen). Na het uitsturen van de laatste brieven op 2 maart 2021 moest er nog een betalingstermijn van 32 dagen worden voorzien (tot en met 5 april 2021). Personen die toen een boetebrief ontvingen, hadden zo de mogelijkheid om tijdig te betalen.

4.4.2 Van batches naar clusters

Voor de verdeling van de groepen en het uitsturen van de brieven maakten we gebruik van een *Cluster Randomized Controlled Trial*. Hierbij gebeurt de randomisatie niet op individuele basis, maar op basis van de clusters waarbinnen individuele eenheden gegroepeerd zijn (Gerber & Green, 2012; Moerbeek, 2006). De keuze voor een geclusterd design is het gevolg van de procedures waarmee GASAM de boetebrief naar individuele overtreders uitstuurt. Dit uitsturen van de boetebrieven gebeurt op basis van batches van boetes

[18] Er werd in overeenstemming met GASAM beslist om een aantal bijzondere dossiers niet mee te nemen in de analyse omdat de betalingstermijn niet op een reguliere manier geëvalueerd kon worden. Het betreft dossiers waarvoor verweer werd ingediend, dossiers in schuldbemiddeling en dossiers gelinkt aan gevangenschap.

die over de periode van een aantal dagen verzameld zijn. Elke batch krijgt één van de vier interventies toegewezen. In 'normale' omstandigheden ontvangt GASAM van de gemeentelijke politie één of tweemaal per week een batch met 50 tot 150 processen-verbaal rond GAS-overtredingen voor fout parkeren of stilstaan. Omwille van COVID-19 bleek de frequentie en de grootte van de batches gedurende de looptijd van ons experiment minder regelmatig.

Na ontvangst van een batch stelt de sanctionerend ambtenaar de boete vast en is GASAM verantwoordelijk voor het uitsturen van de boetebrieven. Dit gebeurt op basis van dezelfde batches aangezien de brieven gelinkt zijn aan het proces-verbaal vanwege de politie. In ons experiment stond GASAM in voor de administratieve opvolging van de boetes. De financiële opvolging gebeurde door de dienst Financiën van Stad Mechelen. Concreet gingen GASAM en de dienst Financiën na of de boetes correct en tijdig waren betaald en bepaalde GASAM of bijkomende maatregelen nodig waren. GASAM verzamelde alle noodzakelijke data voor dit experiment. GASAM kon iedere respondent aan de relevante batch linken, wist welk type brief bij welke batch hoorde, op welke datum de boetebrief verzonden werd, en wanneer de betaling plaatsvond. Daarnaast verzorgde GASAM per dossier ook de informatie over het geslacht en de leeftijd van de respondent die wij gebruikten om de kwaliteit van de randomisatie te beoordelen. Deze data werd geanonimiseerd en ter beschikking gesteld van de onderzoekers.

Tijdens de dataverzameling werden verschillende controle- en monitoringmechanismen ingebouwd om ons ervan te vergewissen dat de implementatieprocedures van het project verliepen zoals gepland. Vanuit GASAM kregen we een maandelijkse update met de grootte en de verdeling van de batches. Verder hebben we evaluatiemomenten ingebouwd na één maand, na drie maanden en na zes maanden. Tijdens afsluiting van de dataverzameling hebben we bovendien per batch manuele steekproeven uitgevoerd om te controleren dat dossiernummers, batchnummers, en de voorziene interventies overeen kwamen.

In het kader van dit veldexperiment waren respondenten zich er niet van bewust dat ze deel uitmaakten van een wetenschappelijke studie. Om praktische en methodologische redenen konden we geen gebruik maken van een vooraf verkregen geïnformeerde toestemming. We hebben er daarom specifiek zorg voor gedragen dat de impact en de lasten van de interventies op het gedrag van deelnemers zo laag mogelijk

bleven. Na afloop van het experiment en de publicatie van dit rapport werden de resultaten van het experiment uitvoerig via de communicatiekanalen van de Stad Mechelen naar respondenten en inwoners van de stad gecommuniceerd.

4.5 RESULTATEN

Gedurende de looptijd van het experiment werden er in totaal n = 4860 overtredingen voor fout parkeren of fout stilstaan in de stad Mechelen vastgesteld. De resultaten van ons onderzoek zijn gebaseerd op de terugbetalingstermijn van de n = 2664 individuen die gedurende de looptijd van het experiment de controlebrief of één van de drie interventiebrieven ontvingen. De individuele brieven zijn acht tot negen keer per batch uitgestuurd, waarbij de batchgroottes varieerden tussen de n = 629 en n = 710 respondenten. De batchgroottes en frequenties van implementatie per batch zijn weergegeven in Tabel 14.

Tabel 14. Aantal respondenten en clustergroottes

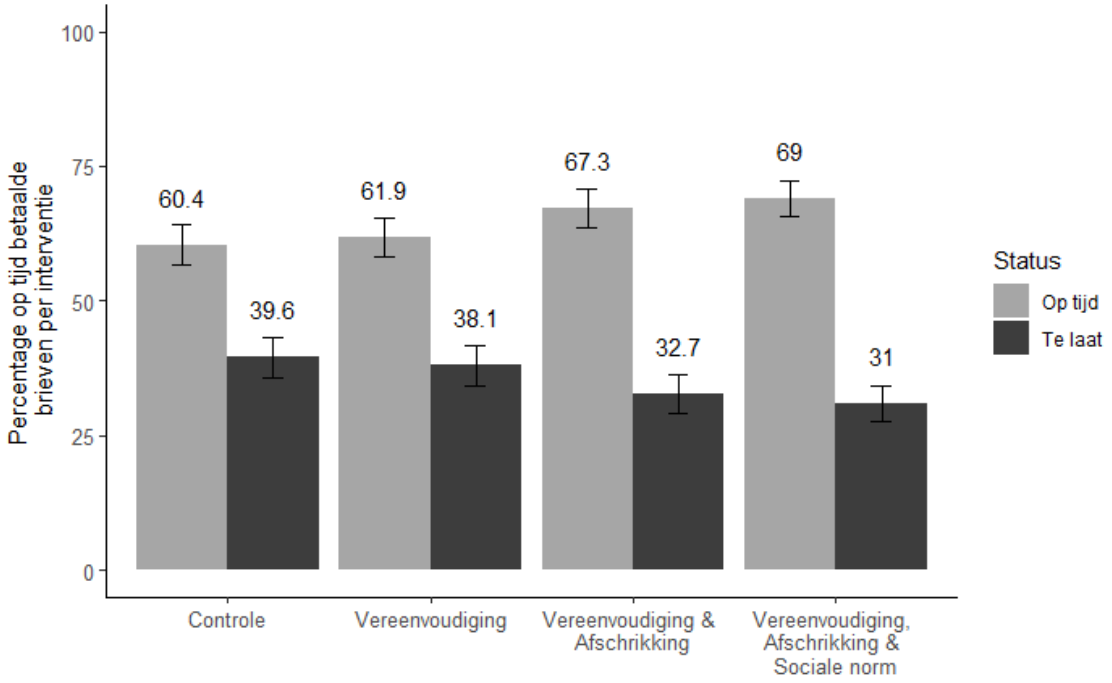
Brief	Interventie	n. Resp.	n. clusters	$\bar{x}$ resp. per cluster
C	Geen interventie (controlegroep)	629	9	70
V	Vereenvoudiging	680	9	76
VA	Vereenvoudiging + afschrikking	645	8	81
VAS	Vereenvoudiging + afschrikking + sociale norm	710	8	89
	Totaal	2664	34	78

4.5.1 Geobserveerde proporties

In Figuur 23 hebben we het geobserveerde percentage van *op tijd* betaalde boetes per type brief afgezet tegen het geobserveerde percentage niet op tijd betaalde boetes. Voor de controlebrief, de brief die voor aanvang van het experiment door GASAM werd gebruikt als boetebrief, ligt het aantal op tijd betaalde boetes op 60.4%. Bij de vereenvoudigde brief steeg het geobserveerde percentage op tijd betaalde brieven tot 61.9%. Dit is een stijging van op tijd betaalde boetes van 1.5 procentpunten ten opzichte van de controle brief. De combinatie van een vereenvoudiging van de brief en een toevoeging van een afschrikkende boodschap resulteerde in een geobserveerde tijdige betalingsgraad van 67.3%. Een stijging van 6.9 procentpunten ten opzichte van de controlebrief. Tot slot, de combinatie van de vereenvoudiging van de

boetebrief, de toevoeging van een afschrikkende boodschap, en de toevoeging van een sociale norm deed het geobserveerde percentage op tijd betaalde boetes verder stijgen naar 69%. Dit komt overeen met een stijging van de geobserveerde betalingsgraad van 8.6 procentpunten.

De toevoeging van communicatieve nudges lijkt er dus toe te leiden dat de geobserveerde betalingsgraad is toegenomen. Om te beoordelen of deze toenames bij toeval tot stand gekomen zijn of daadwerkelijk toe te wijzen zijn aan de interventies, dienden we de significantie van de verschillen in de betalingsgradaties per brief vast te stellen.



Figuur 23. Percentage op tijd betaalde brieven per interventie (op basis van geobserveerde proporties)

4.5.2 Balansverhoudingen

Voordat we tot deze significantietesten over kunnen gaan, gaan we eerst na of onze geclusterde randomisatiemethodologie (herhaaldelijke op cluster gebaseerde implementatie van de controle en interventiebrieven) resulteerde in equivalente subgroepen per interventie. Om het causale effect van interventies vast te kunnen stellen is het essentieel dat de geobserveerde covariaten gelijk verdeeld zijn over de experimentele subgroepen. Mocht dit niet het geval zijn, dan zijn de vastgestelde effecten mogelijk toe

te wijzen aan factoren buiten het experiment. Om de equivalentie tussen de experimentele subgroepen vast te stellen hebben we op twee relevante covariaten, zijnde de leeftijd en het gender van de individuen in de subgroepen, balanstesten uitgevoerd. De resultaten van deze testen zijn weergegeven in Tabel 15 en tonen aan dat de gemiddelde leeftijd en de man/vrouw verhoudingen per subgroep niet significant verschillen. Op basis van deze twee balanstesten concluderen we dat de randomisatie geslaagd is en dat de experimentele subgroepen in voldoende mate equivalent zijn.

Tabel 15. Balanstesten

	Brief C	Brief V	Brief VA	Brief VAS	Balanstest
Gemiddelde leeftijd (in jaren)	43.5	42.7	42.3	42.2	$F(3, 1472) = 1.158$, $p = 0.32$
Man/vrouw verhouding (prop. vrouwen)	.337	.381	.340	.354	$\chi^2(3) = 3.527$, $p = 0.317$

4.5.3 Significantietesten

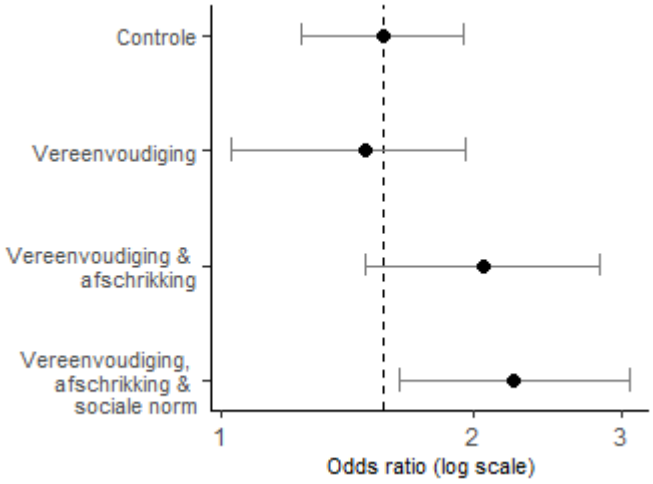
Om te kunnen beoordelen of de interventies significant hebben bijgedragen aan het verbeteren van de tijdige betalingsgraad zijn twee aspecten van belang. Ten eerste is het niet mogelijk om de termijn tot betaling van nog openstaande boetes vast te stellen. De betaling heeft immers nog niet plaatsgevonden. We kijken daarom niet naar de betalingstermijn, maar naar het proportie tijdig betaalde boetes per brief en voorkomen zo dat de nog openstaande boetes de effectschatting van de interventies beïnvloeden. Dit betekent dat wij bestuderen of het proportie tijdig betaalde boetes van één of meer van de drie interventiegroepen significant hoger ligt dan die van de controle brief. We testen daarbij of er significante verschillen bestaan tussen de controle brief en de interventiebrieven in de kans dat iemand op tijd dan wel niet op tijd betaalt.

Ten tweede moeten we de clustering van individuele betalingsgraden in de batches waarin de brieven uitgezonden zijn in acht nemen. Deze clustering betekent dat het effect van de controle- en interventiebrieven op de betalingsgraad van individuen mede afhankelijk is van de cluster waaraan een individuele brief is toegewezen. De kans dat een individu toegewezen aan een controlebrief-batch een interventiebrief zou krijgen was immers nul. Mochten wij deze clustering niet meenemen in onze analyse,

dan overschatten wij de variantie in de data en is er een vergrootte kans dat we significante effecten vaststellen die er eigenlijk niet zijn (type-I-fouten). Om dat te voorkomen is het van belang voor de cluster-gerelateerde variantie te controleren via multi-level benadering, waarin we de cluster-gerelateerde variantie scheiden van de interventie-gerelateerde variantie. Belangrijk daarbij is dat de verdeling van clusters over de controle en interventiebrieven, alsmede de groepsgroottes per cluster in voldoende mate gelijk zijn. De verdeling van clusters over de controle en interventiegroepen en de groepsgroottes per cluster spelen een rol in de efficiëntie van de analyse. Voor experimenten waar deze balans minder of afwezig is, is het relatief moeilijk om significante effecten aan te tonen. Op basis van de bovenstaande overwegingen hebben wij ervoor gekozen de significantie en effectgroottes van de interventiebrieven ten opzichte van de controlebrief op de tijdige betalingsgraad van GAS-boetes te schatten door middel van een multi-level logistische regressieanalyse (e.g., *generalized linear mixed effects regression analysis*). Deze methode stelt ons in staat te testen of er significante verschillen zijn tussen de controle brief en de interventiebrieven in de kans dat een respondent op tijd betaalt of niet op tijd betaalt.[19]

De resultaten van de analyse zijn samengevat in Tabel 16 en grafisch weergegeven in Figuur 24. Figuur 24 toont per brief (controle of interventie) de geschatte kans dat een respondent op tijd betaalt of niet op tijd betaalt (odds ratio) binnen een betrouwbaarheidsinterval van 95%. De kans dat een respondent op tijd betaald versus niet op tijd betaald voor de controle brief is als referentiepunt genomen. Aangezien de betrouwbaarheidsintervallen van de eerste twee interventiebrieven overlappen met die van de controlebrief kunnen wij concluderen dat de eerste twee interventiebrieven (V en VA) de kans op tijdige betaling niet significant verbeteren ten opzichte van de controle brief. Dat geldt niet voor de laatste interventiebrief (VAS). In Figuur 24 tonen we dat respondenten die een vereenvoudigde brief met een afschrikkende en sociale norm boodschap hebben ontvangen significant meer kans hebben om op tijd te betalen vergeleken met respondenten die een controlebrief hebben ontvangen. Gelijktijdig kunnen we op basis van Figuur 24 ook stellen dat de controlebrief significant beter lijkt te functioneren dan 'vereenvoudiging' alleen, maar *niet* significant beter dan een vereenvoudigde brief met een afschrikkende boodschap.

[19] De interventies zijn gecodeerd als dummies, waarbij de controlebrief als referentie optrad en de effecten van interventiebrieven als *differentials* ten opzichte van de referentie *geïnterpreteerd* zijn (op de logit schaal).



Figuur 24. Resultaten interventies in odds ratios (binnen 95% betrouwbaarheidsinterval)

Als we dan naar de cijfers in Tabel 16 kijken, dan zien wij dat de kans dat een individu een boete op tijd betaalt significant hoger ligt dan de kans dat een individu een boete niet op tijd betaalt. Gecontroleerd voor de clustering in de data, is de geschatte kans dat een individu een boete op tijd betaalt versus niet op tijd betaalt 1.56 (OR=1.56, 95% CI: 1.25, 1.95). Respondenten met de controle brief zijn daarmee 56% waarschijnlijker op tijd te betalen dan niet op tijd te betalen. De kans dat een individu die een vereenvoudigde brief (V) ontvingen op tijd betaalt versus niet op tijd betaalt 1.49 (OR=1.49, 95% CI: 1.03, 1.96). Dat is een waarschijnlijkheid van 49%, en hoewel deze waarschijnlijkheid lager ligt dan die voor de controlebrief, is het verschil tussen beiden niet statistisch significant.

Gecontroleerd voor de clustering is de kans dat een individu die een vereenvoudigde brief met een afschrikkende boodschap (VA) ontving op tijd betaalt 2.05 keer hoger dan dat datzelfde individu niet op tijd zou betalen (OR=2.05, 95% CI: 1.48, 2.84). Dat komt overeen met een waarschijnlijkheid van 105%, maar wederom is dit verschil niet significant groter dan bij de controle brief (bij een 95% betrouwbaarheidsinterval). Tot slot is de kans dat respondenten die een vereenvoudigde brief met een afschrikkende boodschap en een sociale norm ontvangen (VAS) op tijd betalen versus niet op tijd betalen 2.24 (OR=2.24, 95% CI: 1.63, 3.08). Het is daarmee 124% waarschijnlijker dat deze mensen op tijd betalen dan

niet op tijd betalen. Daarnaast toont de Wald-test dat individuen die een interventiebrief VAS krijgen een significant hogere kans hebben op tijd te betalen dan individuen die de controle brief krijgen.

Tabel 16. Resultaten interventies in odds ratios (referentie=controle brief)

Interventies	Resultaten in Odds Ratios	
Controle brief (C)	1.56***	(.177)
Vereenvoudiging (V)	.955	(.157)
Vereenvoudiging + afschrikking (VA)	1.32	(.218)
Vereenvoudiging + afschrikking + sociale norm (VAS)	1.43*	(.233)
ICC	.015	

Significantieniveaus: * = p<0.05, ** = p<0.01, *** = P<.001; standaardfouten tussen haakjes.

4.5.4 Experimenteren = lessen leren

Een veldexperiment draait letterlijk om experimenteren in het veld. Dit betekent dat we bepaalde interventies uitproberen in de echte wereld en ingrijpen in het dagelijkse leven van mensen. Dit brengt niet alleen een grote verantwoordelijkheid, maar ook een aantal methodologische en ethische kwesties met zich mee. In het tweede hoofdstuk over de experimentele methode en in het algemene besluit van dit rapport reflecteren we hier uitvoerig over. In dit intermezzo staan we al even stil bij vier praktische en operationele aandachtspunten.

Les 1: Elk eindpunt is een nieuw beginpunt

Bij de meeste veldexperimenten zitten wetenschappelijke onderzoekers en beleidspractici samen rond de tafel, elk met hun eigen methoden, waarden, systemen, verwachtingen en doelen. Nadat deze in een verkennende fase min of meer op elkaar zijn afgestemd, start een intensieve en uitgebreide designfase waarin het experiment wordt ontworpen. Interventies worden uitgedacht en geïmplementeerd, de randomisatieprocedure wordt op punt gezet en de data worden verzameld. Eens alles in de startblokken staat en het experiment begint te lopen, lijkt het of we comfortabel voor een bepaalde periode, en liefst tot de voorziene einddatum, gelanceerd zijn. Schijn bedriegt. Dit eerste eindpunt waarin het design van de studie wordt opgezet is vooral een beginpunt van een nieuwe fase in de studie. Tijdens die fase moet permanent geëvalueerd, teruggekoppeld en bijgestuurd worden om fouten te vermijden en problemen aan te pakken.

Les 2: You win some, you lose some

Oorspronkelijk was het veldexperiment opgezet met vijf groepen en dus vijf interventies. We hadden ook een brief voorzien die enkel de sociale norm zou testen bovenop de vereenvoudiging. Op die manier konden we een afschrikwekkende en een niet-afschrikwekkende boodschap tegenover elkaar plaatsen. Een evaluatie in de laatste maand van de dataverzameling bracht naar boven dat een reeks brieven uit één specifieke batch niet waren verzonden. Respondenten hadden onze vijfde interventiebrief niet ontvangen. Omdat het een dermate grote batch betrof, werden de verschillen qua grootte en frequentie tussen de clusters te groot om deze mee te nemen in de finale analyse. Hierdoor hebben we slechts vier interventies kunnen behouden en kunnen we de sociale norm nudge enkel meten in combinatie met de twee andere nudges en niet afzonderlijk. Deze bijsturing was echter noodzakelijk om de validiteit van het experiment te verzekeren.

Les 3: Werk niet met veronderstellingen, maar test wat werkt

Op basis van vorige studies in een belastingcontext gingen we ervan uit dat het effect van de vereenvoudiging quasi vanzelfsprekend was en dat we bovenop die vereenvoudiging de effecten van andere nudges konden testen. We moesten vaststellen dat die veronderstelling niet klopte. Vereenvoudiging sorteert een klein effect en de kans op een tijdige betaling is zelfs lager dan bij de controlebrief. Een mogelijke verklaring is dat de brief in het verleden al verschillende malen geanalyseerd was met het oog op verbetering, waardoor de marge kleiner was. Daarnaast kan het zijn dat de ontworpen vereenvoudiging het doel voorbijging. We hebben dan wel met verschillende betrokkenen vanuit verschillende invalshoeken en achtergronden naar de boodschap van tekst gekeken, maar mogelijk heeft dit de boodschap minder helder gemaakt voor de burger. In ieder geval hebben we geleerd om voorzichtig te zijn met veronderstellingen in de designfase en eerst (kleinschalig) te testen wat werkt en wat niet.

Les 4: Houd rekening met onverwachte gebeurtenissen

COVID-19 ontziet niets of niemand, ook niet veldexperimenten over parkeerboetes. De timing van ons veldexperiment viel midden in de coronacrisis (juni 2020 - maart 2021). Deze externe gebeurtenis had op minstens twee manieren een invloed op onze planning. Allereerst hebben we de start van het experiment met twee maanden moeten uitstellen. Het operationaliseren en implementeren van de interventies diende op afstand te gebeuren en de permanente beschikbaarheid bij GASAM lag lager waardoor we vertraging opliepen. Ten tweede hebben we de duurtijd van het project met drie maanden moeten verlengen. Omwille

van COVID-19 hadden we de grootte van de steekproef ten opzichte van referentiejaar 2019 fors onderschat. De circulatie van het aantal mensen in een Stad als Mechelen was lager waardoor ook het aantal GAS-boetes voor verkeer een daling kende. Ondanks deze tegenslagen, maar vooral dankzij onze partners in Mechelen zijn we er toch in geslaagd om het veldexperiment in grote lijnen uit te voeren zoals we het op voorhand hadden uitgestippeld.

4.6 CONCLUSIES

De studie toont vooreerst aan dat de drie nudge-interventies (V, VA en VAS) de betalingsgraad van de GAS-boetes in Mechelen deed toenemen ten opzichte van de controlebrief. De controlebrief kwam uit bij een betalingsgraad van 60.4%, wat ongeveer overeenstemde met de gemiddelde betalingsgraad van de voorbij vijf jaar (zie Figuur 18). De vereenvoudiging (+1.5 procentpunten); de vereenvoudiging & afschrikwekkende boodschap (+6.9 procentpunten); en de vereenvoudiging & afschrikwekkende boodschap & sociale norm (+8.6 procentpunten) zorgden er alle drie voor dat meer burgers hun boete op tijd betaalden. Ten tweede blijkt dat enkel de personen die een versimpelde boetebrief kregen mét een afschrikkende boodschap en mét een sociale norm een significant hogere kans hebben op tijd te betalen dan personen die de controle brief krijgen wanneer we ook rekening houden met de grootte en de spreiding van de clusters.

Afzonderlijke nudges in de context van lokale parkeerboetes hebben met andere woorden geen of slechts een bescheiden effect. Een enkele, subtiele nudge via één extra zinnetje aan een brief is niet voldoende om de betalingsgraad te verhogen. Eerder onderzoek wees reeds op de relatief kleine effecten van nudgeboodschappen in belastingbrieven (Antinyan & Asatryan, 2019). Toch kan de context van belastingen niet over de hele lijn gelijk worden gesteld met dit experiment rond parkeerboetes. We werkten in ons onderzoek bijvoorbeeld niet met de aanmaningsbrief zoals gebruikelijk in de belastingcontext, maar met de eerste, administratieve brief zelf die de informatie over de overtreding en de betaling meegeeft. Een mogelijke verklaring voor het beperkte effect van de boodschappen is dat het afzonderlijke, psychologische mechanisme van zulke communicatieve nudges enkel in staat is om een klein, contextafhankelijk effect toe te voegen aan een veel complexer gedragspatroon (Vainre et al., 2020). Ook het feit dat in dit experiment het geanticipeerde effect van de vereenvoudiging uitbleef, vormt mogelijks een belangrijke verklarende factor.

Een laatste vraag die we ons kunnen stellen is wat de financiële effecten van deze nudges zijn. Voor de looptijd van de studie werden er 2664 administratieve sancties uitgeschreven die aan de selectievoorwaarden van het experiment voldeden. Deze sancties besloegen een totaalbedrag van € 184.208. Gedurende het experiment is er voor € 65.540 aan sancties niet op tijd betaald. Op basis van het berekende effect op het betaalgedrag hebben de drie interventiebrieven (in vergelijking met de controlebrief) een geschatte opbrengst meegebracht van € 3.169 aan op tijd betaalde brieven die zonder deze interventies niet op tijd betaald zouden zijn. Voor GASAM en de Stad Mechelen komt dit neer op 1.7% van het totale bedrag of 4.9% van de te laat of niet betaalde bedragen. Voor een lokaal bestuur zijn dit versnelde inkomsten en minder administratiekosten voor de opvolging van te laat betaalde parkeerboetes. Voor de burger betekent dit gemiddeld genomen minder extra administratieve en financiële kosten die gepaard gaan met laattijdige betalingen.

4.7 AANBEVELINGEN

Op basis van de resultaten van dit onderzoek raden we de Stad Mechelen en GASAM aan voor de inning van de GAS-boetes voor foutief stilstaan en parkeren definitief over te schakelen naar de brief die alle nudge-interventies combineert: tegelijkertijd herschikken, afschrikken en op de sociale norm mikken. Het veldexperiment wees uit dat op deze wijze het betaalgedrag significant veranderde. De brief vermindert de administratieve kosten voor lokaal bestuur en burger. De werking ervan kan bovendien de komende jaren zonder experimenteel design verder gemonitord en geëvalueerd worden door te kijken naar de evolutie van de gemiddelde jaarlijkse betalingsgraad. Op termijn zal het dan weliswaar steeds moeilijker worden om het causale verband tussen de gewijzigde brief en het betalingsgedrag aan te tonen.

Verder moeten we voorzichtig zijn met het veralgemenen van deze resultaten naar andere domeinen zoals GAS-boetes voor overlast. Allicht heeft dit domein een eigen dynamiek en gevoeligheden in vergelijking met verkeersovertredingen. Het veldexperiment is met andere woorden contextafhankelijk. Het levert onvoldoende informatie en bewijs voor toepassing op andere domeinen. Het is daarom aan te raden om gelijkaardige veldexperimenten op te zetten voor andere domeinen om te testen welke interventies het meest effectief zijn om gedrag te doen veranderen. Het is niet aangeraden dat een lokaal bestuur zonder

enige reflectie het design van de nieuwe brief voor GAS-boetes in het verkeer overneemt en bijvoorbeeld toepast op GAS-boetes rond overlast.

Daarnaast moeten we opmerken dat we we op basis van deze resultaten niet volledig kunnen achterhalen in welke mate burgers de gehanteerde nudges accepteren. Hoewel we vaststellen dat de (combinatie van) nudges een positief effect hebben op het tijdig betalen van de parkeerboete, hebben we geen zicht op mogelijke neveneffecten zoals bijvoorbeeld het vertrouwen van burgers in de overheid. Als we de nudges die gebruikt werden voor deze studie, met name sociale normen en afschrikwekkende boodschappen, toetsen aan het nudge framework van Hansen & Jespersen (2013), komen we uit bij transparante type 2 nudges, die het keuzeproces van burgers faciliteren en niet manipuleren. Verder weten we dat burgers over het algemeen het gebruik van nudges steunen (Sunstein et al., 2018, 2019). In Vlaanderen bedraagt de gemiddelde publieke acceptatie van nudgemaatregelen bijvoorbeeld 61,7% (Beyst & Rubens, 2018). Al lag de focus in deze studies in de eerste plaats op beleidsdomeinen zoals gezondheid en duurzaamheid. Om een beter beeld te krijgen van de mogelijke neveneffecten van deze specifieke nudges in deze specifieke context is vervolgonderzoek naar publieke acceptatie aangewezen.

Tot slot bestaat een andere, interessante piste erin om niet alleen de administratieve en financiële kosten voor burgers te verminderen maar ook te testen hoe we de administratieve lasten, de zogenaamde *administrative burdens* (Christensen et al., 2020; Moynihan et al., 2015) naar beneden kunnen krijgen. De mogelijkheid om dit soort boetes te betalen via het scannen van een QR code met een smartphone (zoals reeds ingevoerd door de FOD Justitie en de FOD Mobiliteit op federaal niveau) kan ook overwogen worden. Via technologische innovaties kunnen overheden misschien wel een 'vereenvoudiging in het kwadraat' bereiken, en de zogenaamde frictiekosten en uiteindelijk ook de administratieve lasten gevoelig doen dalen.

5 BESLUIT

De simulante integratie van inzichten en methoden uit de gedragswetenschappen in beleidsonderzoek- en praktijk is een fascinerend fenomeen. De opkomst van het gedragskundig beleidsperspectief kunnen we interpreteren als een reactie op de grenzen van een rationeel beleidsperspectief en als appreciatie van een realistischer mensbeeld. Er bestaan verschillende toepassingen (nudges, maar ook boosts), verschillende voordelen (gebruiksvriendelijk en goedkoop) en verschillende kritieken (ineffectief en zelfs manipulatief). De opkomst van het experimentele design kunnen we interpreteren als een reactie op de beperkingen van observationele evaluatiemethoden en als appreciatie van causaliteit. Er bestaan verschillende toepassingen (labo, veld en survey), één uniek voordeel (dé gouden standaard) en verschillende risico's (methodolotry, privacy en discriminatie).

In dit rapport hebben we meer willen doen dan enkel het beschrijven van deze trend richting meer gedragsinzichten en meer experimenten in beleid en bestuurskunde. We hebben gekozen voor de actie en hebben een samenwerking opgezet tussen universiteit en overheid, tussen onderzoekers en practici om beleidsexperimenten op basis van gedragsinzichten te ontwerpen, implementeren en evalueren. Dit resulteerde in twee cases rond verkeersboetes. Samen met de Vlaamse Stichting Verkeerskunde (VSV) hebben we een survey-experiment ontworpen om het effect te testen van rationale en emotionele boodschappen op de intentie van bestuurders om te snel te rijden. Op die manier evalueerden we brieven die in het veld over heel Vlaanderen meegestuurd worden bij een snelheidsboete. Samen met de Stad Mechelen hebben we een geclusterd veldexperiment ontworpen om het effect te testen van verschillende soorten nudges op het al dan niet tijdig betalen van een parkeerboete. Voortbouwend op deze mix van beleidsinzichten, -acties en -ervaringen formuleren we een antwoord op de volgende driedimensionale vraag: Is het ontwerpen van beleidsexperimenten op basis van gedragsinzichten effectief, ethisch en praktisch haalbaar?

De effectiviteitsdimensie hangt samen met de vraag in hoeverre de gedragsinterventies het gedrag van burgers hebben kunnen veranderen. Voor beide experimenten met verkeersboetes vonden we bescheiden effecten of zelfs geen effecten. Zo kon de emotionele boodschap de algemene intentie om te snel te rijden niet significant verminderen, terwijl de rationele boodschap (iets) meer succes kende. Bij het veldexperiment

rond parkeerboetes slaagde uiteindelijk enkel de brief die alle nudge-interventies combineerde erin om het betaalgedrag significant te verbeteren. We concluderen dat bescheiden gedragsinterventies, zoals een emotionele boodschap of sociale norm toevoegen aan een boetebrief, leiden tot bescheiden resultaten. Bovendien blijkt de context waarbinnen gedragsinterventies en nudges getest worden een bepalende factor. Zo is het terugbetalen van achterstallige belastingen aan een federale overheid niet hetzelfde als het terugbetalen van een parkeerboete aan een lokaal bestuur. Deze contextafhankelijkheid kan deels verklaren waarom nudge-interventies die vertrekken vanuit dezelfde theoretische en empirische premissen in verschillende omstandigheden soms tot verschillende resultaten leiden.

De ethische dimensie van beleidsexperimenten op basis van gedragsinzichten situeert zich op twee niveaus: de keuze van psychologische inzichten en de keuze van experimentele methoden. De gedragswetenschappelijk geïnspireerde boodschappen en mechanismen die vervat waren in de boetebrieven van beide experimenten speelden vooral in op bewuste keuzeprocessen (transparant en systeem 2) waardoor de autonomie van de burgers niet fundamenteel werd aangetast. De experimentele designs die we uittekenden voldeden aan de volgende cruciale voorwaarden: het informeren van deelnemers, het garanderen van privacy, het kiezen voor geïnformeerde toestemming of debriefing, en het betrekken van een ethische commissie.

De praktische dimensie omvat de implementatie van inzichten én methoden uit de gedragswetenschappen en hun compatibiliteit met het beleidsproces. We ondervonden dat beide componenten, gedragstheorie en experimentele methode, geen vanzelfsprekende of hapklare ingrediënten zijn die je zonder slag of stoot kan toevoegen en implementeren. De behoorlijk lineaire en ééndimensionale stappenplannen om gedragsinterventies te ontwerpen en via experimentele design te testen - die ook in dit rapport verschijnen - bieden misschien wel handige kapstokken, maar eenmaal je ze confronteert met de beleidsrealiteit getuigen ze ook van een zekere naïviteit. De werkelijkheid is complexer en vereist een zekere dosis flexibiliteit en pragmatisme. Bij het ontwerpen van interventies versmelten de theoretische, gedragswetenschappelijke beginselen met contextuele factoren zoals de praktische kennis en ervaring van beleidspractici en de feedback van burgers. Waardoor uiteindelijk de vraag zich opdringt in hoeverre we eigenlijk nog mogen spreken van "*theory-driven*" interventies (Van Ryzin, 2021).

Hoewel deze drie conclusies misschien niet meteen aanleiding geven tot extra enthousiasme rond de gedragswetenschappelijk en experimentele omslag in bestuur en beleid, kunnen deze bevindingen wel leiden tot een beter begrip en een realistischer beeld van zowel de theorie als de methode achter deze beweging. Juist door de (te) hoge verwachtingen te temperen en te vertrekken vanuit een bescheiden positie van niet (zeker) weten wat werkt, komen we tot de kern van experimenteren: uitzoeken en uitproberen wat wel en niet werkt binnen een specifieke setting. Bovendien heeft dit onderzoek ook drie belangrijke hefbomen kunnen identificeren die de verdere ontwikkeling van beleidsexperimenten kunnen stimuleren en aligneren met de idee van *evidence-informed poliyc making*.

Ten eerste, experimenteren is leren. Het leren vindt plaats op verschillende terreinen en door verschillende betrokkenen. Onderzoekers leren welke hypothesen, (deels) opgetrokken uit theoretische gronden, verworpen of bevestigd kunnen worden en welke wetenschappelijke methoden passen bij welk soort onderzoeksvragen Beleidspractici leren rekening houden met het gedrag van burgers en kennis over dit gedrag mee te nemen in het beleidsproces. Daarnaast komen ze ook op basis van welke methoden en designs beleidsmaatregelen geëvalueerd kunnen worden en komen ze te weten welke beleidsacties werken en welke niet. Soms komen beleidsonderzoekers en beleidspractici er zelfs achter waarom iets al dan niet werkt. Bij het survey-experiment rond rijsnelheid konden we dankzij het gebruik van de *Protection Motivation Theory* achterhalen dat de emotionele brief effect miste omdat ze bestuurders onvoldoende concrete tips en handvaten aanreikte om zich in de toekomst aan de gepaste rijsnelheid te houden . Bij het veldexperiment rond parkeerboetes werd duidelijk dat een enkele nudge niet volstaat, maar dat een combinatie van verschillende nudges meer potentieel heeft. Deze resultaten bieden informatie om een gefundeerde beslissing te nemen over het schrappen, behouden, aanpassen of uitrollen van bepaalde beleidsacties.

Ten tweede, experimenteren is kalibreren. Het komt erop aan om de verschillende componenten van het experimentele design – de proefpersonen, de interventie, de randomisatieprocessen, de dataverzameling, de analyse en de evaluatie – zo goed mogelijk op de geijkte theoretische en methodologische standaarden en op elkaar af te stemmen. Zeker voor beleidsexperimenten in het veld geldt echter dat hoe hoger het realisme (de *fieldness*), hoe meer oncontroleerbare factoren en onvoorziene beperkingen er opduiken. Zo

zorgden vaste procedures inzake datatransfer en ICT beheer ervoor dat we in Mechelen opteerden voor geclusterde in plaats van een individuele randomisatie. En ook bij het survey-experiment waren we tot op zekere hoogte gebonden aan de toon en de inhoud van de algemene campagne rond rijsnelheid en de op voorhand vastgelegde contouren voor de evaluatie van deze campagne. Dit kalibreren beperkt zich bovendien niet tot de ontwerpfase, maar blijft cruciaal tijdens de implementatiefase en de evaluatiefase van het experiment. Zelden verloopt een experiment immers helemaal zoals het op voorhand gepland was en op papier stond. Op de schouders van onderzoekers en practici rust de taak om doorheen dit grillige traject de integriteit, kwaliteit en validiteit van het experiment te waarborgen.

Ten derde, experimenteren is co-creëren of samenwerken (en zo van elkaar leren). Ook het samenwerken vindt plaats op verschillende niveaus en tussen verschillende betrokkenen. Het onderzoeksteam dient verschillende kennisdomeinen en methodologische vaardigheden te combineren, en is dus gebaat bij een polyvalente en multidisciplinaire samenstelling. In dit geval was er kennis nodig over gedragsverandering, beleidsinstrumenten, beleidsprocessen, experimenteel design, statistische analyse en projectmanagement. Ook vanuit de beleidskant werden zowel bij VSV als bij de Stad Mechelen verschillende profielen, diensten en partners betrokken. Zo werden er bij het veldexperiment met de Stad Mechelen mensen uit de interlokale vereniging GASAM, de dienst Financiën, de dienst Innovatie en de dienst Marketing & Communicatie ingeschakeld en kregen we ondersteuning het team Gedragsinzichten (BUZA) en het team Heerlijk Helder van de Vlaamse overheid. Bij het survey-experiment werkten we via VSV ook samen met een communicatiebureau en een marktonderzoeksbureau. Boven alles bleek vooral de brug tussen onderzoekers en practici bijzonder waardevol en vruchtbaar. Door als volwaardige partners samen te werken wordt snel duidelijk wat onderzoekers en practici voor elkaar kunnen betekenen bij dit soort projecten en ontstaat de mogelijkheid om van elkaar te leren en stappen vooruit te zetten op weg naar meer en beter onderbouwd beleid.

6 REFERENTIES

Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis and Prevention*, *38*(2), 215–224. https://doi.org/10.1016/j.aap.2005.07.004

Afif, Z., Islan, W., Calvo-Gonzalez, O., & Dalton, A. G. (2018). *Behavioral Science Around the World: Profiles of 10 Countries*.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211. https://doi.org/10.4135/9781446249215.n22

Al-Ubaydli, O., Lee, M. S., List, J. A., 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. https://doi.org/10.1017/bpp.2020.17

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095. https://doi.org/10.1016/j.jpubeco.2011.03.003

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, *104*(10), 3003–3037. https://doi.org/10.1257/aer.104.10.3003

Allewaert, T., & De Sutter, S. (2010). *Gemeentelijke administratieve sancties. Balans 1999-2009*. die Keure.

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, *1*, 323–338.

Alm, J., Jackson, B. R., & McKee, M. (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics*, *93*(3–4), 392–402. https://doi.org/10.1016/j.jpubeco.2008.10.007

Anderson, D. M., & Edwards, B. C. (2015). Unfulfilled Promise: Laboratory experiments in public management research. *Public Management Review*, *17*(10), 1518–1542. https://doi.org/10.1080/14719037.2014.943272

Antinyan, A., & Asatryan, Z. (2019). Nudging for Tax Compliance: A Meta-Analysis. *SSRN Electronic Journal*, *November*. https://doi.org/10.2139/ssrn.3500744

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *Leadership Quarterly*, *28*(1), 5–21. https://doi.org/10.1016/j.leaqua.2017.01.006

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly*, *21*(6), 1086–1120. https://doi.org/10.1016/j.leaqua.2010.10.010

Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial. *Criminology*, *50*(1), 27–69. https://doi.org/10.1111/j.1745-9125.2011.00256.x

Armitage, C. J., & Conner, M. (2001). Efficacy of the Theory of Planned Behaviour : A Meta-Analytic Review. *British Journal of Social Psychology*, *40*, 471–499.

Arno, A., & Thomas, S. (2016). The efficacy of nudge theory strategies in influencing adult dietary behaviour: A systematic review and meta-analysis. *BMC Public Health*, *16*(1), 1–11. https://doi.org/10.1186/s12889-016-3272-x

Baekgaard, M., Baethge, C., Blom-Hansen, J., Dunlop, C. A., Esteve, M., Jakobsen, M., Kisida, B., Marvel, J., Moseley, A., Serritzlew, S., Stewart, P., Thomsen, M. K., & Wolf, P. J. (2015). Conducting Experiments in Public Management Research: A Practical Guide. *International Public Management Journal*, *18*(2), 323–342. https://doi.org/10.1080/10967494.2015.1024905

Ball, S., & Head, B. W. (2020). Behavioural insights teams in practice: nudge missions and methods on trial. *Policy & Politics*, *xx*(xx), 1–16. https://doi.org/10.1332/030557320x15840777045205

Banerjee, S., & John, P. (2021). Nudge plus: incorporating reflection into behavioral public policy. *Behavioural Public Policy*, 1–16. https://doi.org/10.1017/bpp.2021.6

Bantjes, J., Plessis, S. du, Jansen, A., Siebrits, K., & Slabbert, P. (2021). Motorists' perceptions of factors that

influence payment of speeding fines in Cape Town, South Africa: application of the Theory of Planned Behaviour. *South African Journal of Psychology*. https://doi.org/10.1177/00812463211011746
Battaglio, R. P., Belardinelli, P., Bellé, N., & Cantarelli, P. (2019). Behavioral Public Administration ad fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations. *Public Administration Review*, *79*(3), 304–320. https://doi.org/10.1111/puar.12994
Bauwens, T., Pleysier, S., De Pauw, E., & Van den Broeck, T. (2016). De nieuwe GAS-wet. Zuurstof of stikstof voor steden en gemeenten? *Orde van de dag*, *73*, 4–10.
Bekkers, V., Fenger, M., & Scholten, P. (2017). *Public Policy in Action. Perspectives on the policy process*. Edward Elgar Publishing Limited.
Bellé, N. (2013). Experimental Evidence on the Relationship between Public Service Motivation and Job Performance. *Public Administration Review*, *73*(1), 143–153. https://doi.org/10.1111/j.1540-6210.2012.02621.x
Bellé, N. (2014). Leading to make a difference: A field experiment on the performance effects of transformational leadership, perceived social impact, and public service motivation. *Journal of Public Administration Research and Theory*, *24*(1), 109–136. https://doi.org/10.1093/jopart/mut033
Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should Governments Invest More in Nudging? *Psychological Science*, *28*(8), 1041–1055. https://doi.org/10.1177/0956797617702501
Bertelli, A. M., & Riccucci, N. M. (2020). What Is Behavioral Public Administration Good for? *Public Administration Review*, *00*, 1–5. https://doi.org/10.1111/puar.13283
Beyst, V., & Rubens, K. (2018). Wordt 'nudging' in het beleid aanvaard in Vlaanderen? *Vlaams Tijdschrift voor Overheidsmanagement*, *23*(4), 55–67.
Biesta, G. (2007). Why "what works" won't work: evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *57*(1), 1–22. https://doi.org/10.1111/j.1741-5446.2006.00241.x
BIN-NL. (2018). *Gedragstechnieken voor brieven en e-mails*. https://binnl.nl/kennis/publicaties/gedragstechnieken+voor+brieven+en+e-mails/default.aspx
BIN NL. (2019). *Hoe Is Gedrag Te Onderzoeken*. https://binnl.nl/default.aspx
BIT. (2014). *EAST: Four Simple Ways to Apply Behavioural Insights*. BIT UK.
BIT UK. (2017). *The Behavioural Insighst Team. Annual Update 2016-2017*.
Blom-Hansen, J., Morton, R., & Serritzlew, S. (2015). Experiments in Public Management Research. *International Public Management Journal*, *18*(2), 151–170. https://doi.org/10.1080/10967494.2015.1024904
Bloom, H. S. (2005). *Learning from Social Experiments*. Russell Sage.
Botzem, S. (2019). The governance of behavioural taxation: moralization and the new models of tax collection. In H. Strassheim & S. Beck (Red.), *Handbook of behavioural change and public policy* (pp. 272–283). Edward Elgar Publishing Limited.
Bouwman, R., & Grimmelikhuijsen, S. (2016). *Experimental public administration from 1992 to 2014 : a systematic literature review and ways forward Citing information : Bouwman , R ., & Grimmelikhuijsen , S . G . 2016 . " Experimental public ways forward ", International Journal of Public Sector Man*. *March*.
Bozeman, B. (1992). *Experimental Design in Public Policy and Management Research : Introduction*. *2*(4), 440–442.
Brockmeyer, A., Smith, S., Hernandez, M., & Kettle, S. (2019). Casting a wider tax net: Experimental evidence from Costa Rica. *American Economic Journal: Economic Policy*, *11*(3), 55–87. https://doi.org/10.1257/pol.20160589
Brown, S. R., & Melamed, L. E. (1990). *Experimental Design and Analysis (Quantitative Applications in the Social Sciences, Nº 74)*. Sage Publications.
Bunn, F., Collier, T., Frost, C., Ker, K., Roberts, I., & Wentz, R. (2003). Traffic calming for the prevention of road traffic injuries: Systematic review and meta-analysis. *Injury Prevention*, *9*(3), 200–204.

https://doi.org/10.1136/ip.9.3.200
Burden, B. C., Canon, D. T., Mayer, K. R., & Moynihan, D. P. (2012). The Effect of Administrative Burden on Bureaucratic Perception of Policies: Evidence from Election Administration. *Public Administration Review*, *72*(5), 741–751. https://doi.org/10.1111/j.1540-6210.2012.02600.x
Cairney, P., Oliver, K., & Wellstead, A. (2016). To Bridge the Divide between Evidence and Policy: Reduce Ambiguity as Much as Uncertainty. *Public Administration Review*, *76*(3), 399–402. https://doi.org/10.1111/puar.12555
Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, *24*, 409–429.
Chen, T., Sze, N. N., Saxena, S., Pinjari, A. R., Bhat, C. R., & Bai, L. (2020). Evaluation of penalty and enforcement strategies to combat speeding offences among professional drivers: A Hong Kong stated preference experiment. *Accident Analysis and Prevention*, *135*(November 2019). https://doi.org/10.1016/j.aap.2019.105366
Chirico, M., Inman, R., Loeffler, C., MacDonald, J., & Sieg, H. (2019). Deterring Property Tax Delinquency in Philadelphia. *National Tax Journal*, *72*(3), 479–506. https://doi.org/10.17310/ntj.2019.3.01
Christensen, J., Aarøe, L., Baekgaard, M., Herd, P., & Moynihan, D. P. (2020). Human Capital and Administrative Burden: The Role of Cognitive Resources in Citizen-State Interactions. *Public Administration Review*, *80*(1), 127–136. https://doi.org/10.1111/puar.13134
Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science*, *12*(4), 105–109. https://doi.org/10.1111/1467-8721.01242
Colson, G., Corrigan, J. R., Grebitus, C., Loureiro, M. L., & Rousu, M. C. (2016). Which Deceptive Practices, If Any, Should Be Allowed in Experimental Economics Research? Results from Surveys of Applied Experimental Economists and Students. *American Journal of Agricultural Economics*, *98*(2), 610–621. https://doi.org/10.1093/ajae/aav067
Congdon, W. J., & Shankar, M. (2018). The Role of Behavioral Economics in Evidence-Based Policymaking. *Annals of the American Academy of Political and Social Science*, *678*(1), 81–92. https://doi.org/10.1177/0002716218766268
Conner, M., Lawton, R., Parker, D., Chorlton, K., Manstead, A. S. R., & Stradling, S. (2007). Application of the theory of planned behaviour to the prediction of objectively assessed breaking of posted speed limits. *British Journal of Psychology*, *98*(3), 429–453. https://doi.org/10.1348/000712606X133597
Cops, D., Put, J., & Pleysier, S. (2012). De GAS-wetgeving als instrument voor de aanpak van overlast. Bedenkingen bij en suggesties voor een integraal en onderbouwd overlastbeleid. *Panopticon: Tijdschrift voor Strafrecht, Criminologie en Forensisch Welzijnswerk*, *33*(6), 552–563.
Cranor, T., Goldin, J., Homono, T., & Moore, L. (2020). Communicating Tax Penalties to Delinquent Taxpayers: Evidence from a Field Experiment. *National Tax Journal*, *73*, 331–360.
Czibor, E., Jimenez-Gomez, D., & List, J. A. (2019). The Dozen Things Experimental Economists Should Do (More of). *Southern Economic Journal*, *86*(2), 371–432. https://doi.org/10.1002/soej.12392
De Ceunynck, T., Pelssers, B., & Daniels, S. (2020). *Gevleugelde oversteekplaatsen op 2x2 wegen – Evaluatiestudie aan de hand van gedrags- en conflictobservatie*.
De Ceunynck, T., Pelssers, B., Daniels, S., Temmerman, P., & Martensen, H. (2020). *3D voetgangersoversteekplaatsen – Evaluatiestudie aan de hand van gedrags- en conflictobservatie en snelheidsmetingen*.
De Francesco, F., & Pattyn, V. (2021). Policy evaluation styles. In M. Howlett & J. Tosun (Red.), *The Routledge Handbook of Policy Styles*. Routledge.
de Graaf, A., Sanders, J., & Hoeken, H. (2016). Characteristics of narrative interventions and health effects: A review of the content, form, and context of narratives in health-related narrative persuasion research. *Review of Communication Research*, *4*, 81–131. https://doi.org/10.12840/issn.2255-4165.2016.04.01.011
De Neve, J. E., Imbert, C., Spinnewijn, J., Tsankova, T., & Luts, M. (2021). How to improve tax compliance?

Evidence from population-wide experiments in Belgium. *Journal of Political Economy*, *129*(5), 1425–1463. https://doi.org/10.1086/713096
De Peuter, B., De Smedt, J., & Bouckaert, G. (2007). *Handleiding Beleidsevaluatie. Deel 1: Evaluatiedesign en -management*.
de Ridder, D., & Tummers, L. (2019). *Nudging. Makkelijke oplossingen voor moeilijke keuzes*. Prometheus.
De Schepper, T., Van Heddeghem, K., & Verbeek, M. (2013). *Gemeentelijke administratieve sancties. De bestuurlijke aanpak van overlast*. Politeia.
Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, *210*(December 2017), 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005
Delhomme, P., De Dobbeleer, W., Forward, S., & Simões, A. (2009). *Manual for designing, implementing, and evaluating road safety communication campaigns*.
Della Vigna, S., & Linos, E. (2020). RCTs to Scale: Comprehensive Evidence from Two Nudge Units. *NBER Working Paper Series*, *27594*(March), 73. http://www.nber.org/papers/w27594
Desmidt, S., & Meyfroodt, K. (2020). How Does Public Disclosure of Performance Information Affect Politicians' Attitudes towards Effort Allocation? Evidence from a Survey Experiment. *Journal of Public Administration Research and Theory*, 756–772. https://doi.org/10.1093/jopart/muaa054
Devroe, E. (2012). *A swelling culture of control? De genese en toepassing van de wet op de gemeentelijke administratieve sancties*. Maklu.
Druckman, J. N., Green, D., Kuklinski, J. H., & Lupia, A. (2011). *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press.
du Plessis, S., Hartig, B., Jansen, A., & Siebrits, K. (2020). Improving payment of traffic fines with financial incentives: Discounts vs. penalties. *Transportation Research Part F: Traffic Psychology and Behaviour*, *74*(March 2021), 298–306. https://doi.org/10.1016/j.trf.2020.08.019
Duflo, E., Glennerster, R., & Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. *Handbook of Development Economics*, *4*, 3895–3962.
Dunning, T. (2012). *Natural Experiments in the Social Science: A Design-Based Approach*. Cambridge University Press.
Dye, T. R. (2017). *Understanding public policy (fifteenth edition)*. Pearson.
Eden, D. (2017). Field Experiments in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*(February), 91–122. https://doi.org/10.1146/annurev-orgpsych-041015-062400
Eeckhout, B. (2013, mei 29). Stop de verGASsing, nu het nog kan. *De Morgen*. https://www.demorgen.be/nieuws/stop-de-vergassing-nu-het-nog-kan~b4c61c7d/
Einfeld, C. (2019). Nudge and evidence based policy: fertile ground. *Evidence & Policy*, *15*(4), 509–524.
Elliott, M. A., Armitage, C., & Baughan, C. (2003). Drivers' compliance with speed limits: an application of the theory of planned behavior. *Journal of Applied Psychology*, *88*, 964–972.
Esmark, A. (2019). Communicative governance at work: how choice architects nudge citizens towards health, wealth and happiness in the information age. *Public Management Review*, *21*(1), 138–158. https://doi.org/10.1080/14719037.2018.1473476
Ewert, B. (2020). Moving beyond the obsession with nudging individual behaviour: Towards a broader understanding of Behavioural Public Policy. *Public Policy and Administration*, *35*(3), 337–360. https://doi.org/10.1177/0952076719889090
Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, *326*(5952), 535–538. https://doi.org/10.1126/science.1168244
Feitsma, J. (2019a). *Inside the Behavioural State*. Eleven International Publishing.
Feitsma, J. (2019b). Meters maken: gedragsexperts in het lokaal bestuur. *Bestuurswetenschappen*, *73*(1), 5–29. https://doi.org/10.5553/bw/016571942019073001002
Feitsma, J. (2020). 'Rationalized incrementalism'. How behavior experts in government negotiate institutional logics. *Critical Policy Studies*, *14*(2), 156–173. https://doi.org/10.1080/19460171.2018.1557067

Ferraro, P. J., & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics*, *95*(1), 64–73. https://doi.org/10.1162/REST_a_00344
Field, A., & Hole, G. (2002). *How to Design and Report Experiments*. Sage.
Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministery of Agriculture of Great Britain*, *33*, 503–513.
Floyd, D., Prentice-Dunn, S., & Rogers, R. (2000). A Meta-Analysis of Research on Protection Motivation Theory. *Journal of Applied Psychology*, *30*(2), 407–429.
Frechette, G. R., & Schotter, A. (2015). *Handbook of Experimental Economic Methodology*. Oxford University Press.
Friedman, D. (2010). *Experimental Methods: A Primer for Economics*. Princeton University Press.
Friedman, L. M., Furberg, C. D., & DeMets, D. L. (2010). *Fundamentels of Clinical Trials*. Springer.
GASAM. (2019). *Gemeentelijke Administratieve Sancties Arrondissement Mechelen (GASAM) - Jaarverslag 2018*.
GASAM. (2020). *GASAM Jaarverslag 2019*.
GASAM. (2021). *Interne dataset GASAM 2015-2020*.
George, B., Baekgaard, M., Decramer, A., Audenaert, M., & Goeminne, S. (2020). Institutional isomorphism, negativity bias and performance information use by politicians: A survey experiment. *Public Administration*, *98*(1), 14–28. https://doi.org/10.1111/padm.12390
Gerber, A. S., & Green, D. P. (2012). *Field Experiments. Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc.
Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The Foundations of Adaptive Behavior*. Oxford Scholarship Online.
Glendon, A. I., & Walker, B. L. (2013). Can anti-speeding messages based on protection motivation theory influence reported speeding intentions? *Accident Analysis and Prevention*, *57*, 67–79. https://doi.org/10.1016/j.aap.2013.04.004
Glennerster, R., & Takavarasha, K. (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton University Press.
Goldenbeld, C., Groot-Mesken, J., & Temürhan, M. (2017). *Nudging van rijsnelheid via Dick Bruna-borden: een veldexperiment. De effecten op werkelijk gereden snelheden in vijf gemeenten onderzocht. . R-2017-11*.
Graf, R. (2019). Nudging before the nudge? Behavioural traffic safety regulation and the rise of behavioural economics. In H. Strassheim & S. Beck (Red.), *Handbook of behavioural change and public policy* (pp. 23–37). Edward Elgar Publishing.
Greenberg, D., Linksz, D., & Mandell, M. (2003). *Social Experimentation and Public Policymaking*. Urban Institute Press.
Grimmelikhuijsen, S. G., & Meijer, A. J. (2014). Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. *Journal of Public Administration Research and Theory*, *24*(1), 137–157. https://doi.org/10.1093/jopart/mus048
Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review*, *77*(1), 45–56. https://doi.org/10.1111/puar.12609
Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge Versus Boost: How Coherent are Policy and Theory? *Minds and Machines*, *26*(1–2), 149–183. https://doi.org/10.1007/s11023-015-9367-9
Haglund, M., & Åberg, L. (2000). Speed choice in relation to speed limit and influences from other drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *3*(1), 39–51. https://doi.org/10.1016/S1369-8478(00)00014-0
Hallsworth, M. (2014). The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy*, *30*(4), 658–679. https://doi.org/10.1093/oxrep/gru034
Hallsworth, M., Chadborn, T., Sallis, A., Sanders, M., Berry, D., Greaves, F., Clements, L., & Davies, S. C. (2016).

Provision of social norm feedback to high prescribers of antibiotics in general practice: A pragmatic national randomised controlled trial. *The Lancet*, *387*(10029), 1743–1752. https://doi.org/10.1016/S0140-6736(16)00215-4
Hallsworth, M., & Kirkman, E. (2020). *Behavioral insights*. MIT Press.
Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14–31. https://doi.org/10.1016/j.jpubeco.2017.02.003
Halpern, D. (2015). *Inside the nudge unit: How small changes can make a big difference*. WH Allen.
Halpern, D., & Mason, D. (2015). Radical Incrementalism. *Evaluation*, *21*(2), 143–149. https://doi.org/10.1177/1356389015578895
Hansen, J. A., & Tummers, L. (2020). A Systematic Review of Field Experiments in Public Administration. *Public Administration Review*, *80*(6), 921–931. https://doi.org/10.1111/puar.13181
Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, *4*(1), 3–28. https://doi.org/10.1017/s1867299x00002762
Harrison, G., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *XLII*(December), 1009–1055. https://doi.org/10.4337/9781788110563.00013
Hassan, S., & Wright, B. E. (2020). The Behavioral Public Administration Movement: A Critical Reflection. *Public Administration Review*, *80*(1), 163–167. https://doi.org/10.1111/puar.13130
Hauser, O. P., Linos, E., & Rogers, T. (2017). Innovation with field experiments: Studying organizational behaviors in actual organizations. *Research in Organizational Behavior*, *37*, 185–198. https://doi.org/10.1016/j.riob.2017.10.004
Haynes, L., Green, D., Gallagher, R., John, P., & Torgerson, D. (2013). Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages. *Journal of Policy Analysis and Management*, *32*(4), 718–730. https://doi.org/10.1002/pam
Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). *Test , Learn , Adapt : Randomised Controlled Trials*. BIT UK.
Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science*, *12*(6), 973–986. https://doi.org/10.1177/1745691617702496
Hinyard, L. J., & Kreuter, M. W. (2007). Using narrative communication as a tool for health behavior change: A conceptual, theoretical, and empirical overview. *Health Education and Behavior*, *34*(5), 777–792. https://doi.org/10.1177/1090198106291963
Holocher, S., & Holte, H. (2019). *Speeding. ESRA2 Thematic report Nr. 2. ESRA project (E-Survey of Road users' Attitudes)*.
Hössinger, R., & Berger, W. J. (2012). Stated response to increased enforcement density and penalty size for speeding and driving unbelted. *Accident Analysis and Prevention*, *49*, 501–511. https://doi.org/10.1016/j.aap.2012.03.023
House of Lords Science and Technology Select Comittee. (2011). *Behaviour Change. Science and Technology Select Comitee.. 2nd report of session 2010-2012*. Authority of the House of Lords.
Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying Public Policy. Policy Cycles & Policy Subsystems*. Oxford University Press.
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*(February), 47–58. https://doi.org/10.1016/j.socec.2019.03.005
IBZ. (2021). *GAS-verslag 2021: wat is belangrijk om te weten?* https://www.besafe.be/nl/nieuws/gas-verslag-2021-wat-is-belangrijk-om-te-weten
James, O., Jilke, S., & Van Ryzin, G. (2017a). Causal Inference and Experiments. In O. James, S. Jilke, & G. Van Ryzin (Red.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 59–88).

Cambridge University Press.
James, O., Jilke, S., & Van Ryzin, G. G. (2017b). *Experiments in Public Management Research. Challenges and Contributions*. Cambridge University Press.
James, O., John, P., & Moseley, A. (2017). Field Experiments in Public Management. In O. James, S. Jilke, & G. van Ryzin (Red.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 89–116). Cambridge University Press.
James, O., & Moseley, A. (2014). Does performance information about public services affect citizens' perceptions, satisfaction, and voice behaviour? Field experiments with absolute and relative performance information. *Public Administration*, *92*(2), 493–511. https://doi.org/10.1111/padm.12066
Jensen, N. H., Reuss, L. F., & Rasmussen, S. (2018). Increasing Public Debt Collection With Nudging: Results of Two Natural Fields Experiments. *Scandinavian Journal of Public Administration*, *22*(4), 45–64.
Jilke, S., Van Dooren, W., & Rys, S. (2018). Discrimination and administrative burden in public service markets: Does a public-private difference exist? *Journal of Public Administration Research and Theory*, *28*(3), 423–439. https://doi.org/10.1093/jopart/muy009
Jilke, S., & Van Ryzin, G. (2017). Survey experiments for Public Management Research. In O. James, S. Jilke, & G. Van Ryzin (Red.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 117–138). Cambridge University Press.
John, P. (2014). Policy entrepreneurship in UK central government: The behavioural insights team and the use of randomized controlled trials. *Public Policy and Administration*, *29*(3), 257–267. https://doi.org/10.1177/0952076713509297
John, P. (2017a). All tools are informational now: *Rethinking policy and politics*, 183–202. https://doi.org/10.2307/j.ctt1t895hw.15
John, P. (2017b). Changing How Government Works: The Transformative Potential of an Experimental Public Management. In O. James, S. Jilke, & G. Van Ryzin (Red.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 476–494). Cambridge University Press.
John, P. (2017c). *Field experiments in Political Science and Public Policy. Practical lessons in Design and Delivery*. Routledge.
John, P. (2018). *How Far to Nudge? Assessing Behavioural Public Policy*. Edward Elgar Publishing Limited.
John, P. (2020). Improving Causal Claims in Public Policy through Randomized Designs in the Field. *Journal of Comparative Policy Analysis: Research and Practice*, *00*(00), 1–17. https://doi.org/10.1080/13876988.2020.1782751
John, P., & Blume, T. (2018). How best to nudge taxpayers? The impact of message simplification and descriptive social norms on payment rates in a central London local authority. *Journal of Behavioral Public Administration*, *1*(1). https://doi.org/10.30636/jbpa.11.10
John, P., & Stoker, G. (2019). Rethinking the role of experts and expertise in behavioural public policy. *Policy and politics*, *47*(2), 209–226. https://doi.org/10.1332/030557319X15526371698257
Jones, R. A. (1985). *Research methods in the social and behavioral sciences*. Sinauer Associates.
Jones, R., Pykett, J., & Whitehead, M. (2013). *Changing behaviours: On the rise of the psychological state*. Edward Elgar Publishing.
Kahneman, D. (2012). *Ons feilbare denken*. Business Contact.
Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.
Kahneman, D., & Tversky, A. (1979). Prospect Theory : An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292. http://www.jstor.org/stable/1914185%5Cnhttp://www.jstor.org/%5Cnhttp://www.jstor.org/action/showPublisher?publisherCode=econosoc.%5Cnhttp://www.jstor.org
Keiser, L. R., & Miller, S. M. (2020). Does Administrative Burden Influence Public Support for Government Programs? Evidence from a Survey Experiment. *Public Administration Review*, *80*(1), 137–150.

https://doi.org/10.1111/puar.13133
Kergoat, M., Delhomme, P., & Meyer, T. (2017). Appraisal of speed-enforcement warning messages among young drivers: Influence of automatic versus human speed enforcement in a known or unknown location. *Transportation Research Part F: Traffic Psychology and Behaviour*, *46*, 177–194. https://doi.org/10.1016/j.trf.2017.01.005
Kettle, S., Hernandez, M., Ruda, S., & Sanders, M. (2016). *Behavioral Interventions in Tax Compliance: Evidence from Guatemala. World Bank Policy Research Working Paper No. 7690* (Nummer June). https://doi.org/10.1596/1813-9450-7690
Kinder, D. R. (2011). Campbell's ghost. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Red.), *Cambridge Handbook of Experimental Political Science* (pp. 525–530). Cambridge University Press.
Kirchler, E. (2007). *The Economic Psychology of Tax Behaviour*. Cambridge University Press.
Kirk, R. E. (2012). *Experimental Design: Procedures for the Behavioural Sciences*. Sage Publications.
Knoepfel, P., Larrue, C., Varone, F., & Hill, M. (2011). *Public Policy Analysis*. The Policy Press.
Kok, G., Gottlieb, N. H., Peters, G. J. Y., Mullen, P. D., Parcel, G. S., Ruiter, R. A. C., Fernández, M. E., Markham, C., & Bartholomew, L. K. (2016). A taxonomy of behaviour change methods: an Intervention Mapping approach. *Health Psychology Review*, *10*(3), 297–312. https://doi.org/10.1080/17437199.2015.1077155
Krosnick, J. (1999). Survey Research. *Annual Review of Psychology*, *50*, 537–567. https://doi.org/10.1097/ALN.0000000000002436
Langbeen, T., & Thomas, M. (2020). *Opinie: GAS-boetes in corona-tijden. Zal COVID-19 onze maatschappij "vergassen"?*
Larkin, C., Sanders, M., Andresen, I., & Algate, F. (2019). Testing local descriptive norms and salience of enforcement action: A field experiment to increase tax collection. *Journal of Behavioral Public Administration*, *2*(1), 1–11. https://doi.org/10.30636/jbpa.21.54
Legrand, T. (2012). Overseas and over here: Policy transfer and evidence-based policy-making. *Policy Studies*, *33*(4), 329–348. https://doi.org/10.1080/01442872.2012.695945
Lehner, M., Mont, O., & Heiskanen, E. (2016). Nudging – A promising tool for sustainable consumption behaviour? *Journal of Cleaner Production*, *134*. https://doi.org/10.1016/j.jclepro.2015.11.086
Leong, C., & Howlett, M. (2020). Theorizing the behavioral state: Resolving the theory-practice paradox of policy sciences. *Public Policy and Administration*. https://doi.org/10.1177/0952076720977588
Lepenies, R., & Malecka, R. (2019). The ethics of behavioural public policy. In A. Lever & A. Paoma (Red.), *The Routledge handbook of ethics and public policy* (pp. 513–525). Routledge.
Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World ? *Journal of Economic Perspectives*, *21*(2), 153–174.
Lewis, I., Watson, B., Tay, R., & White, K. M. (2007). The role of fear appeals in improving driver safety: A review of the effectiveness of fear-arousing (threat) appeals in road safety advertising. *International Journal of Behavioral Consultation and Therapy*, *3*(2), 203–222. https://doi.org/10.1037/h0100799
Lewis, I., Watson, B., & White, K. M. (2008). An examination of message-relevant affect in road safety messages: Should road safety advertisements aim to make us feel good or bad? *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(6), 403–417. https://doi.org/10.1016/j.trf.2008.03.003
Lewis, I., Watson, B., & White, K. M. (2009). What do we really know about designing and evaluating road safety advertising? Current knowledge and future challenges. In R. Grzebieta & D. McTiernan (Red.), *Proceedings of the 2009 Australasian Road Safety Research, Policing and Education and the 2009 Intelligent Speed Adaption (ISA) Conference* (pp. 733–746). Roads and Traffic Authority, New South Wales.
Li, H., & Van Ryzin, G. (2017). A Systematic Review of Experimental Studies in Public Management Journals. In J. Oliver, S. Jilke, & G. Van Ryzin (Red.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 20–36). Cambridge University Press.

Lodge, M., & Wegrich, K. (2016). The Rationality Paradox of Nudge: Rational Tools of Government in a World of Bounded Rationality. *Law and Policy*, *38*(3), 250–267. https://doi.org/10.1111/lapo.12056
Loer, K. (2019). The enzymatic effect of behavioural sciences – What about policy-maker's expectations? In H. Strassheim & S. Beck (Red.), *Handbook of behavioural change and public policy* (pp. 180–194). Edward Elgar Publishing.
Luca, M., & Bazerman, M. H. (2020). *The power of experiments: Decision making in a data-driven world*. MIT Press.
Luts, M., & Van Roy, M. (2018). Nudging in een belastingcontext. Hoe FOD Financiën dankzij gedragsinzichten burgers aanzet om sneller te betalen. *Vlaams Tijdschrift voor Overheidsmanagement*, *23*(4), 43–53.
MacKay, K., & Quigley, M. (2018). Exacerbating Inequalities? Health Policy and the Behavioural Sciences. *Health Care Analysis*, *26*(4), 380–397. https://doi.org/10.1007/s10728-018-0357-y
Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology*, *19*(5), 469–479. https://doi.org/10.1016/0022-1031(83)90023-9
Madi, M. A. (2020). *The Dark Side of Nudges*. Routledge.
Margetts, H. Z. (2011). Experiments for public management research. *Public Management Review*, *13*(2), 189–208. https://doi.org/10.1080/14719037.2010.532970
McClendon, G. H. (2012). Ethics of Using Public Officials As Field Experiment Subjects. *Newsletter of the APSA Experimental Section*, *3*(1), 13–20.
Mcgraw, K. M., & Scholz, J. T. (1991). Appeals to Civic Virtue versus Attention to Self-Interest : Effects on Tax Compliance. *Law & Society Review*, *25*(3), 471–498.
Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, *6*(1). https://doi.org/10.1186/1748-5908-6-42
Migchelbrink, K., & Van de Walle, S. (2019). When Will Public Officials Listen? A Vignette Experiment on the Effects of Input Legitimacy on Public Officials' Willingness to Use Public Participation. *Public Administration Review*, *80*, 271–280. https://doi.org/10.1111/puar.13138
Milliron, V., & Toy, D. (1988). Tax compliance: An investigation of key features. *The Journal of the American Taxation Association*, *9*(1), 84–104.
Moerbeek, M. (2006). Cluster Randomized Trials: Design and Analysis. In H. Pham (Red.), *Springer Handbook of Engineering Statistics* (pp. 705–718). Springer London. https://doi.org/10.1007/978-1-84628-288-1_39
Mols, F., Haslam, S. A., Jetten, J., & Steffens, N. K. (2015). Why a nudge is not enough: A social identity critique of governance by stealth. *European Journal of Political Research*, *54*(1), 81–98. https://doi.org/10.1111/1475-6765.12073
Morton, R. B., & Williams, K. (2010). *Experimental Political Science: From Nature to the Lab*. Cambridge University Press.
Moynihan, D. (2018). A great schism approaching? Towards a micro and macro public administration. *Journal of Behavioral Public Administration*, *1*(1). https://doi.org/10.30636/jbpa.11.15
Moynihan, D., Herd, P., & Harvey, H. (2015). Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions. *Journal of Public Administration Research and Theory*, *25*(1), 43–69. https://doi.org/10.1093/jopart/muu009
Moynihan, D. P. (2013). Does Public Service Motivation Lead to Budget Maximization? Evidence from an Experiment. *International Public Management Journal*, *16*(2), 179–196. https://doi.org/10.1080/10967494.2013.817236
Mukherjee, I., & Giest, S. (2020). Behavioural Insights Teams (BITs) and Policy Change: An Exploration of Impact, Location, and Temporality of Policy Advice. *Administration and Society*. https://doi.org/10.1177/0095399720918315

Nagatsu, M., & Favereau, J. (2020). Two Strands of Field Experiments in Economics: A Historical-Methodological Analysis. *Philosophy of the Social Sciences*, *50*(1), 45–77. https://doi.org/10.1177/0048393119890393
Nielsen, P. A., & Baekgaard, M. (2015). Performance information, blame avoidance, and politicians' attitudes to spending and reform: Evidence from an experiment. *Journal of Public Administration Research and Theory*, *25*(2), 545–569. https://doi.org/10.1093/jopart/mut051
Olsen, A. L. (2015). "Simon Said," We didn't jump. *Public Administration Review*, *75*(2), 325–326. https://doi.org/10.1111/puar.12330
Olsen, A. L. (2017). Human Interest or Hard Numbers? Experiments on Citizens' Selection, Exposure, and Recall of Performance Information. *Public Administration Review*, *77*(3), 408–420. https://doi.org/10.1111/puar.12638
Özkan, T., Lajunen, T., & Summala, H. (2006). Driver Behaviour Questionnaire: A follow-up study. *Accident Analysis and Prevention*, *38*(2), 386–395. https://doi.org/10.1016/j.aap.2005.10.012
Paris, H., & Van den Broucke, S. (2008). Measuring cognitive determinants of speeding: An application of the theory of planned behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(3), 168–180. https://doi.org/10.1016/j.trf.2007.09.002
Parker, D., Stradling, S. G., & Manstead, A. S. R. (1996). Modifying Beliefs and Attitudes to Exceeding the Speed Limit: An Intervention Study Based on the Theory of Planned Behavior. *Journal of Applied Social Psychology*, *26*, 1–19.
Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205.
Pilkington, P., & Kinra, S. (2005). Effectiveness of speed cameras in preventing road traffic collisions and related casualties: systematic review. *British Medical Journal*, *330*(7487), 331–334.
Podsakoff, P. M., & Podsakoff, N. P. (2019). Experimental designs in management and leadership research: Strengths, limitations, and recommendations for improving publishability. *Leadership Quarterly*, *30*(1), 11–33. https://doi.org/10.1016/j.leaqua.2018.11.002
Pomeranz, D., & Vila-Belda, J. (2019). Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities. *Annual Review of Economics*, *11*, 755–781. https://doi.org/10.1146/annurev-economics-080218-030312
Raymaekers, P. (2020). De gedragsfactor: het nut van nudging en gedragsinzichten in het coronabeleid. *Vlaams Tijdschrift voor Overheidsmanagement*, *3*.
Raymaekers, P., & Brans, M. (2020). Nudging in perspectief. De verbreding van gedragsinzichten in beleid. *Beleidsonderzoek Online*. https://doi.org/10.5553/BO/221335502020000004001
Raymaekers, P., Fobé, E., van Acker, W., & Brans, M. (2019). *Nudging in perspectief. Een verkennend kader voor de toepassing van gedragsinzichten in beleid*.
Raymaekers, P., & Migchelbrink, K. (2021). To nudge or not to nudge? De mogelijkheden van gedragsinzichten en nudging als lokaal beleidsinstrument. *Impuls*, *2*, 6–17.
Reason, J., Manstead, A., Stephen, S., Baxter, J., & Campbell, K. (1990). Errors and violations on the roads: A real distinction? *Ergonomics*, *33*(10–11), 1315–1332. https://doi.org/10.1080/00140139008925335
Riccucci, N. M., Van Ryzin, G. G., & Lavena, C. F. (2014). Representative bureaucracy in policing: Does it increase perceived legitimacy? *Journal of Public Administration Research and Theory*, *24*(3), 537–551. https://doi.org/10.1093/jopart/muu006
Riccucci, N. M., Van Ryzin, G. G., & Li, H. (2016). Representative Bureaucracy and the Willingness to Coproduce: An Experimental Study. *Public Administration Review*, *76*(1), 121–130. https://doi.org/10.1111/puar.12401
Rubaltelli, E., Manicardi, D., Orsini, F., Mulatti, C., Rossi, R., & Lotto, L. (2021). How to nudge drivers to reduce speed: The case of the left-digit effect. *Transportation Research Part F: Traffic Psychology and Behaviour*, *78*, 259–266. https://doi.org/10.1016/j.trf.2021.02.018

Sanders, M., Snijders, V., & Hallsworth, M. (2018). Behavioural science and policy: where are we now and where are we going? *Behavioural Public Policy*, *2*(2), 144–167. https://doi.org/10.1017/bpp.2018.17
Schmidt, S. (2009). Shall We Really Do It Again? The Powerful Concept of Replication Is Neglected in the Social Sciences. *Review of General Psychology*, *13*(2), 90–100. https://doi.org/10.1037/a0015108
Schram, F., & Lievens, J. (2015). *Gemeentelijke administratieve sancties: een antwoord op overlast?* Vanden Broele.
Schultz, P. W. (1999). Basic and Applied Social Psychology Changing Behavior With Normative Feedback Interventions : A Field Experiment on Curbside Recycling Changing Behavior W ith Normative Feedback Interventions : A Field Experim ent on Curbside Recycling. *Basic and Applied Social Psycology*, *21*(1), 25–36. https://doi.org/10.1207/s15324834basp2101
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designsfor Generalized Causal Interference*. Houghton Mifflin Company.
Sherman, L. W., & Weisburd, D. (1995). General detterent effects of police patrol in crime hot spots: a randomized, controlled trial. *Justice Quarterly*, *12*, 755–781.
Sievert, M. (2021). A replication of "Representative bureaucracy and the willingness to coproduce". *Public Administration*, *February*, 1–17. https://doi.org/10.1111/padm.12743
Simon, H. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, *69*(1), 99–118.
Slemrod, J. B. (2018). Tax Compliance and Enforcement: New Research and Its Policy Implications. In *Ross School of Business Paper* (Nummer No. 1302). https://doi.org/10.2139/ssrn.2726077
Slemrod, J., Blumenthal, M., & Christian, C. (2001). Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, *79*(3), 455–483. https://doi.org/10.1016/S0047-2727(99)00107-3
Slemrod, J., & Weber, C. (2012). Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, *19*(1), 25–53. https://doi.org/10.1007/s10797-011-9181-0
Spiliopoulos, L., & Hertwig, R. (2020). A map of ecologically rational heuristics for uncertain strategic worlds. *Psychological Review*, *127*(2), 245–280.
Srinivasan, T. (1973). Tax Evasion: A Model. *Journal of Public Administration Research and Theory*, *2*, 339–346.
Stradling, S. (2000). Drivers who speed. *Impact*, *9*, 38–41.
Strassheim, H. (2019). Behavioural mechanisms and public policy design: Preventing failures in behavioural public policy. *Public Policy and Administration*, 1–18. https://doi.org/10.1177/0952076719827062
Strassheim, H., & Beck, S. (2019). *Handbook of Behavioural Change and Public Policy*. Edward Elgar Publishing.
Sunstein, C. R. (2020). *Behavioral Science and Public Policy*. Harvard University. https://doi.org/10.1017/9781108973144
Sunstein, C. R., Reisch, L. A., & Kaiser, M. (2019). Trusting nudges? Lessons from an international survey. *Journal of European Public Policy*, *26*(10). https://doi.org/10.1080/13501763.2018.1531912
Sunstein, C. R., Reisch, L. A., & Rauber, J. (2018). A worldwide consensus on nudging? Not quite, but almost. *Regulation and Governance*, *12*(1). https://doi.org/10.1111/rego.12161
Sunstein, C. R., & Thaler, R. H. (2003). Libertarian Paternalism Is not an Oxymoron. *University of Chicago Law Review*, *70*(4), 1159–1202. https://doi.org/10.2307/1600573
Taylor, M., Lynam, D., & Baruya, A. (2000). *The effects of drivers' speed on the frequency of road accidents (TRL Report 421).*
Tepe, M., & Prokop, C. (2017). Laboratory Experiments: Their Potential for Public Management Research. In *Experiments in Public Management Research: Challenges and Contributions* (pp. 139–164). Cambridge University Press.
Thaler, R. H. (2015). *Misbehaving. The Making of Behavioral Economics*. W.W. Norton & Company.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth and Happiness*. Penguin.
Thaler, R. H., & Sunstein, C. R. (2018). *Nudge. Naar betere beslissingen over gezondheid, geluk en welvaart.* Uitgeverij Business Contact.
Thau, M., Mikkelsen, M. F., Hjortskov, M., & Pedersen, M. J. (2021). Question order bias revisited: A split-ballot experiment on satisfaction with public services among experienced and professional users. *Public Administration*, *99*(1), 189–204. https://doi.org/10.1111/padm.12688
The Behavioural Insights Team. (2014). *Evaluation of the DECC/John Lewis energy labelling trial*.
The Behavioural Insights Team. (2016). *The Behavioural Insights TeamUpdate Report 2015-16*. http://38r8om2xjhhl25mw24492dir.wpengine.netdna-cdn.com/wp-content/uploads/2016/09/BIT_Update_Report_2015-16-.pdf
Thomke, S. H. (2020). *Experimentation works: The surprising power of business experiments*. MIT Press.
Torgerson, C. J. (2001). The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, *49*(3), 316–328. https://doi.org/10.1111/1467-8527.t01-1-00178
Torgler, B. (2007). *Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis*. Edward Elgar Publishing.
Tummers, L. (2019a). Gedragen gedragsverandering. *Bestuurskunde*, *29*(2), 83–90. https://doi.org/10.5553/bk/092733872019029002012
Tummers, L. (2019b). Public Policy and Behavior Change. *Public Administration Review*, *79*(6), 925–930. https://doi.org/10.1111/puar.13109
Vainre, M., Aaben, L., Paulus, A., Koppel, H., Tammsaar, H., Telve, K., Koppel, K., Beilmann, K., & Uusberg, A. (2020). Nudging towards tax compliance: A fieldwork-informed randomised controlled trial. *Journal of Behavioral Public Administration*, *3*(1), 1–10. https://doi.org/10.30636/jbpa.31.84
Van de Walle, S. (2017). The Experimental Turn in Public Management: How Methodological Preferences Drive Substantive Choices. In O. James, S. Jilke, & G. Van Ryzin (Red.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 461–475). Cambridge University Press.
Van den Berghe, W., & Pelssers, B. (2020). *Themadossier nr. 9 - Snelheid en te snel rijden.*
Van der Knaap, P., Pattyn, V., & Hanemaayer, D. (2020). *Beleidsevaluatie in theorie en praktijk*. Boom bestuurskunde.
Van Ryzin, G. (2021). Nudging and Muddling through. *Perspectives on Public Management and Governance*, 1–7.
van Thiel, S. (2015). *Bestuurskundig onderzoek. Een methodologische inleiding*. Uitgeverij Coutinho.
Vandenbroele, J., Slabbinck, H., Van Kerckhove, A., & Vermeir, I. (2018). Curbing portion size effects by adding smaller portions at the point of purchase. *Food Quality and Preference*, *64*(January 2019), 82–87. https://doi.org/10.1016/j.foodqual.2017.10.015
Vedung, E. (1998). Policy instruments: typologies and theories. In E. Vedung (Red.), *Carrots, sticks and sermons: policy instruments and their evaluation* (pp. 21–58). Transaction Publishers.
Versavel, E., & Kouwenhoven, J. (2020). De nudginaanpak van de no-show-problematiek bij de Directie-Generaal Personen met een Handicap. *Vlaams Tijdschrift voor Overheidsmanagement*, *73*(3), 91–100.
Walker, R. M., Lee, J. M., & James, O. (2017). Replication of Experimental Research: Implications for the Study of Public Management. In O. James, S. Jilke, & G. Van Ryzin (Red.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 439–460). Cambridge University Press.
Warner, H. W., & Åberg, L. (2006). Drivers' decision to speed: A study inspired by the theory of planned behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *9*(6), 427–433. https://doi.org/10.1016/j.trf.2006.03.004
Warner, H. W., & Forward, S. (2016). The effectiveness of road safety interventions using three different messages: Emotional, factual or a combination of both messages. *Transportation Research Part F: Traffic Psychology and Behaviour*, *36*, 25–34. https://doi.org/10.1016/j.trf.2015.11.002

Wayenberg, E. (2020). *Bouwen aan beleid. Het proces van de overheid*. Owl Press.
Whitehead, M., Jones, R., & Pykett, J. (2020). Questioning post-political perspectives on the psychological state: Behavioural public policy in the Netherlands. *Environment and Planning C: Politics and Space*, *38*(2), 214–232. https://doi.org/10.1177/2399654419867711
Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. *Communication Monograhps*, *59*(4), 329–349.
Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education and Behavior*, *27*(5), 591–615. https://doi.org/10.1177/109019810002700506
Wundersitz, L., Hutchinson, T., & Woolley, J. (2010). *Best practice in road safety mass media campaigns: A literature review* (Vol. 5, Nummer 2010).
Yannis, G., Laiou, A., Theofilatos, A., & Dragomanovits, A. (2016). *Speeding. ESRA thematic report no. 1.*
Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly.