

# Improving Public Policy Using Behavioural Understanding

The Behavioural Insights Group Rotterdam

Malte Dewies



Cover and Layout: Malte Dewies  
Printing: Ipskamp Printing

Copyright © 2022 Malte Dewies

All rights reserved. No part of this dissertation may be reproduced or transmitted in any form or by any means without prior written permission from the author

# Improving Public Policy Using Behavioural Understanding

**The Behavioural Insights Group Rotterdam**

Openbaar beleid verbeteren door inzicht in gedrag

De Behavioural Insights Group Rotterdam

## **Thesis**

to obtain the degree of Doctor from the  
Erasmus University Rotterdam  
by command of the rector magnificus

**Prof.dr. A.L. Bredenoord**

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on  
Thursday, 6 October 2022 at 13.00 hrs  
by

**Malte Dewies**

born in Mönchengladbach (Germany)

## **Doctoral Committee**

### **Promotors:**

Prof.dr. S. Denktaş

Prof.dr. K.I.M. Rohde

### **Other members:**

Prof.dr. M.A. Buijzen

Prof.dr. R. van Baaren

Prof.dr. L.G. Tummers

### **Copromotor:**

Dr. I. Merkelbach





# Contents

<b>Introduction</b>		<b>1</b>
<b>Part 1</b>		
<b>1</b>	Applying Behavioural Insights to Public Policy: An Example From Rotterdam	<b>9</b>
<b>Part 2</b>		
<b>2</b>	Call First, Pay Later: Stimulating Debtors to Contact Their Creditors Improves Debt Collection in the Context of Financial Scarcity	<b>29</b>
<b>3</b>	Nudging is Ineffective When Attitudes Are Unsupportive: An Example From a Natural Field Experiment	<b>45</b>
<b>4</b>	Committing to Keep Clean: Nudging Complements Standard Policy Measure To Reduce Illegal Urban Garbage Disposal in a Neighborhood With High Levels of Social Cohesion	<b>63</b>
<b>5</b>	No Pyro, No Party: Social Factors, Deliberate Choices, and Shared Fan Culture Determine the Use of Illegal Fireworks in a Soccer Stadium	<b>77</b>
<b>6</b>	Comprehensive Evaluation of the Behavioural Insights Group Rotterdam	<b>95</b>
<b>Part 3</b>		
<b>7</b>	Do You Consider Human Behaviour to Be Fixed or Malleable? Your Answer Can Influence Your Policy Choices	<b>123</b>
<b>8</b>	Strategies for Behavioural Insights Teams	<b>133</b>
<b>Discussion</b>		<b>151</b>
<b>Appendices</b>	Summary	<b>165</b>
	Nederlandse samenvatting (Summary in Dutch)	<b>173</b>
	References	<b>183</b>
	Curriculum Vitae	<b>213</b>
	Portfolio	<b>215</b>





# Introduction

Typically, governments assume that humans act rationally (Jones, 2017; Lunn, 2012; Sutherland, 2018). They believe that humans attend to all relevant information, that they process that information in a non-biased way to make correct judgements, and that they act in accordance with these objective judgements. However, as both personal experience and research tell (Kahneman, 2011), humans often fail to act rationally. For instance, they postpone decisions about retirement savings because they implicitly see greater value in spending money now than spending it in an uncertain future; they prefer a positively framed 80% fat-free yoghurt over a negatively framed 20% fat yoghurt; they are more likely to reuse their hotel towel only after reading that others do the same. If humans were fully rational, they would act differently, and recent research therefore challenges the conception of rationality as the major predictor of behaviour. The rationality assumption that governments base their policy on is thus at least partially wrong (e.g., Jones, 2017; Whitehead et al., 2017). How can governments improve the way they reflect on and address behaviour for better health, wealth, and well-being? This dissertation aims to provide an answer to that question.

To capitalise on a more realistic understanding of behaviour, governments need to acquire and adopt such an understanding, think along with it, and act accordingly. Put differently, such understanding needs to be integrated into the formulation and implementation of public policies. In the following, I will first describe what a realistic behavioural understanding entails and what are public policies. Thereafter, I will elaborate on their integration for what has been called behavioural public policy (Strassheim & Beck, 2019), behavioural governance (Gofen et al., 2021), or simply “the behavioural state” (Feitsma, 2019b).

A more *realistic understanding of human behaviour* here means being able to explain why behaviours occur (i.e., knowing behavioural determinants) and knowing how to change them (i.e., behaviour change techniques). This understanding, also called behavioural insights, is based on multidisciplinary research results (Lourenço et al., 2016). These disciplines include but are not limited to psychology, behavioural economics, neurosciences, sociology, and anthropology. They can be subsumed under the umbrella term “behavioural sciences” (Lepenes et al., 2018). There are many definitions of *public policy*. In the simplest form, public policy is “anything a government chooses to do or not to do” (Dye, 1972). More elaborate definitions typically refer to connected (in)actions and decisions of government actors to achieve a selection of goals by specific means in a purposive manner (e.g., Anderson, 1984; Jenkins, 1978).

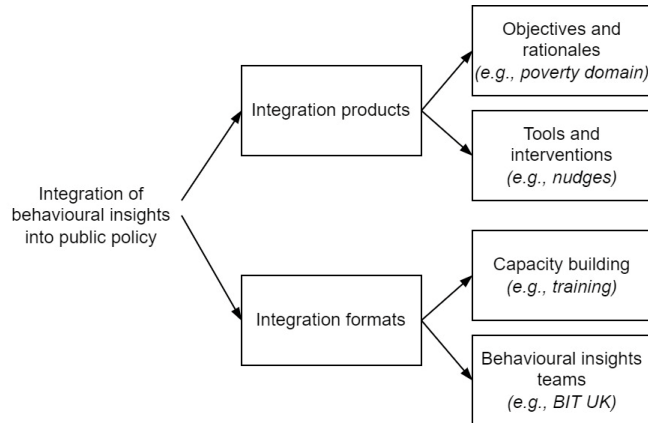
## Integrating Behavioural Insights Into Public Policy

The importance of integrating behavioural insights and public policy has been highlighted by Schneider and Ingram (1990) who claimed already more than 30 years ago that “public policy almost always attempts to get people to do things that they might not otherwise”. Public policy thus aims for behavioural change for which behavioural insights form an essential knowledge base. Indeed, recent evidence shows that integrating behavioural insights into public policies tends to produce more efficient and effective policies (Benartzi et al., 2017; DellaVigna & Linos, 2020; Hummel & Maedche, 2019). Research into the integration of behavioural insights into public policies has explored both

integration products (“What behavioural insights can be integrated into public policies?”) and integration formats (“How to integrate behavioural insights into public policies?”; Figure I.1).

**Figure I.1**

Integration of behavioural insights into public policy



### ***Integration Products***

Related to integration *products*, nudges are arguably the most popular example. Nudges are small and seemingly irrelevant changes in choice contexts that reliably change behaviour, often by exploiting behavioural automatisms such as heuristics and biases (Thaler & Sunstein, 2008). To illustrate, governments have increased the rate of organ donors by changing from opt-in to opt-out defaults (Johnson & Goldstein, 2003). Other examples of integration products include small interventions to foster competencies for making better decisions (“boosts”; Hertwig & Grüne-Yanoff, 2017) and stimuli that create space to deliberately think about and debate solutions (“thinks”; John et al., 2009). These examples concern concrete policy tools or behavioural interventions to be employed instrumentally, but they do not refer directly to goals and rationales underlying public policy. Yet, integrating behavioural insights into public policy can also have as a product that such goals and rationales change based on a particular behavioural understanding. This can be shown using poverty as an example. Here, recent behavioural insights question the common focus on alleviating individual deficiencies (e.g., low financial literacy) and personal attributions of responsibility for poverty (e.g., “one gets what one deserves”; Curchin, 2016; Tiemeijer, 2016; Walker, 2012). Rather, behavioural insights point to the importance of contextual features that (re-)produce poverty and indebtedness (Gennetian & Shafir, 2015) and, therefore, argue for changing these features rather than individuals. In general, integration products have been and continue to be investigated extensively.

### ***Integration Formats***

Integration *formats* often remain hidden and have been investigated much less (Ball & Head, 2021; Gofen et al., 2021). Generally, two formats have been outlined: the first is building up behavioural

insights related capacities within government professionals and the second are centralised teams that use and promote behavioural insights within governments (Lunn, 2012; Sunstein, 2014). In the “age of behavioural science” (Sunstein, 2016), the former option appears somewhat inevitable (Jones, 2017) and behavioural insights continue to be integrated into professional curricula and training content (e.g., Baggio et al., 2021). Moreover, multiple popular science books disseminating behavioural thinking (e.g., Ariely, 2008; Cialdini, 1984; Kahneman, 2011) have been worldwide bestsellers and multiple reports from governments and various (inter)national organisations have advocated and showcased the use of behavioural insights (e.g., United Nations Innovation Network, 2021; World Bank, 2015).

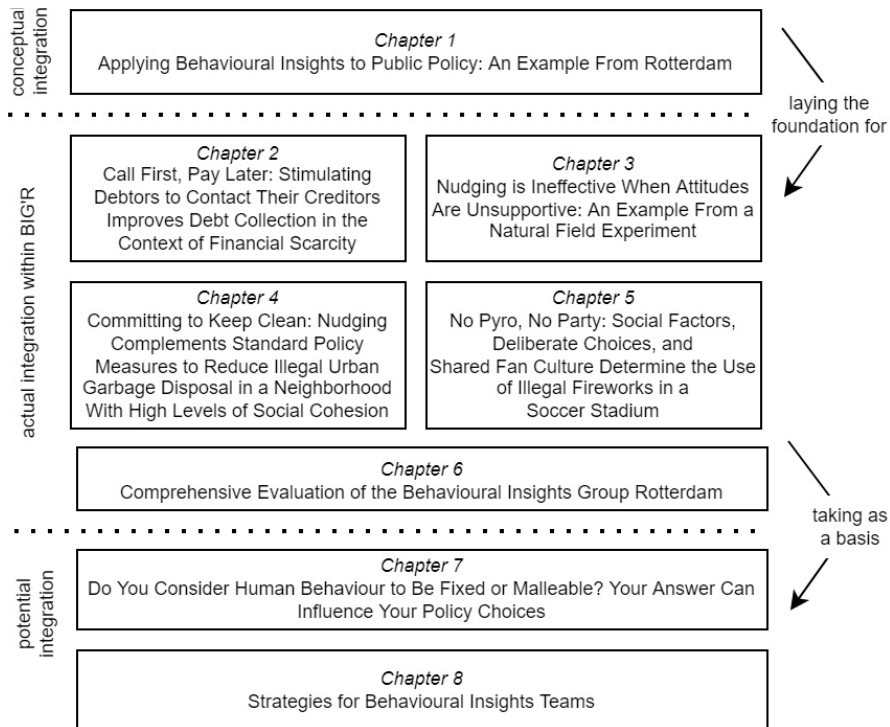
The latter option of integration formats, employing so-called behavioural insights teams (BITs), gained popularity quickly after the successes of the first BIT, set up at the United Kingdom Cabinet Office in 2010, became visible (e.g., John, 2019; 2014). By now, several hundred BITs exist within various organisations (e.g., Afif et al., 2019; Manning et al., 2020; OECD, 2017). Yet, it is often unclear how BITs operate (Ball & Head, 2021; Gofen et al., 2021). This complicates the set-up of BITs and projecting their benefits and related hurdles. The literature and experience from practice suggest though that most BITs rely on scientific methods and an analytic agenda for changing behaviour (Einfeld, 2019; Feitsma, 2019a; 2020; John, 2013). This mainly refers to the use of experimental methods to test behavioural interventions, mostly nudges, in the field (Haynes et al., 2012). BITs can thus be associated with general trends favouring experimentation and a focus on evidence-based practices within public policy (e.g., John, 2014; Leigh, 2018).

To advance and deepen the knowledge about BITs, this dissertation aims to answer the following research question. *How can behavioural insights be integrated centrally into public policy?* For answering this question, I investigated one BIT, the Behavioural Insights Group Rotterdam (BIG'R). This group was the BIT of the municipality of Rotterdam in the Netherlands and existed for four years. It can be described as a “public sector innovation lab” (McGann et al., 2018) for investigating the integration of behavioural insights into public policies in practice.

## Overview

To answer the research question, I adopt a three-stepped approach (see Figure 1.2): First, I describe a conceptual plan for the integration. Second, I describe and evaluate the actual integration within BIG'R. Third, I examine general integration strategies to be applied in the future taking actual integration as a point of departure.

Generally, this research uses a pragmatist approach to research (Johnson & Onwuegbuzie, 2004; Morgan, 2007; 2014) combining both quantitative and qualitative methods (e.g., interviews, surveys, observations). As some of my co-authors and I were active members of BIG'R, this research can be described as “embedded research” which refers to locating researchers within non-academic organisations (Ward et al., 2021a; Ward et al., 2021b). In the following, I describe the three steps and their chapters in more detail.

**Figure I.2****Dissertation overview****Part 1: Conceptual Integration**

In **Chapter 1**, I describe the BIG'R approach for integrating behavioural insights into public policy at both a conceptual and a practical level. The chapter lays the foundation for the following chapters in describing how BIG'R operated. Specifically, it describes what inputs BIG'R used and how it combined them. It also touches upon the outputs that BIG'R anticipated.

**Part 2: Actual Integration**

In part 2 of this dissertation, I report four examples of the actual integration of behavioural insights into public policy and, additionally, a comprehensive evaluation of actual integration within BIG'R. The four examples concern concrete policy issues that BIG'R addressed aiming to develop interventions and advice for the municipality of Rotterdam.

I report the first example in **Chapter 2**. For this example, the BIG'R approach was used to improve debt collection among debtors experiencing financial scarcity. Such debtors often struggle to repay debts in full and immediately. Therefore, BIG'R adapted a repayment letter to encourage those debtors to contact the municipality and agree on payment by instalments. Results show that the adapted repayment letter led to more debtors responding to the letter and agreeing on payment by instalments.

I report a second example of actual integration in **Chapter 3**. Two nudges were developed to increase compliance with an information security policy in a workplace setting. The nudges aimed to counteract forgetting to comply, and to exploit social norms to increase motivation to comply. However, the nudges did not change compliance. Survey results imply that the nudges were ineffective because they failed to affect attitudes about the security policy and because they led to reactance among those targeted by the nudges.

The third example of actual integration in **Chapter 4** concerns nudges used to complement and improve a standard policy approach aiming to reduce illegal littering. This approach was door-to-door canvassing informing residents about rules and regulations for garbage disposal. The nudges added a commitment device to this standard approach for residents to keep their neighbourhood clean. Field testing provided support for the effectiveness of the nudges.

In **Chapter 4**, I report another example of actual integration using the BIG'R approach. It is different from the other examples because it does not report the field testing of an intervention. Rather, it concerns an investigation into the factors that made football supporters use illegal fireworks in a football stadium during matches. Using interviews with supporters and professionals, this chapter explores policy directions to contain the use of fireworks. Results indicate that from a behavioural perspective, stimulating a dialogue between supporters and various stakeholders to seek compromises for containing the use of fireworks seems more promising than the use of repressive measures.

The four examples are sorted with increasing complexity and novelty. The first example uses nudges and tests their effectiveness in a way similar to the trials of other BITs (e.g., Hallsworth, 2017). The second example not only tests the effectiveness of two nudges, but also evaluates how the nudges were perceived. For the third example, nudges are used in a novel way to complement standard policy measures. The fourth example illustrates the application of behavioural insights beyond nudges and for policy formulation early during the policy cycle.

I evaluate actual integration within BIG'R in general (i.e., across individual examples) in **Chapter 5**. This chapter focusses on BIG'R as a whole and over its total period of existence to investigate the inputs of BIG'R, how they were used for what activities, with what outputs, and to which effects.

### ***Part 3: Potential Integration***

In **Chapter 7**, I investigate how assumptions about behaviour influence policy preferences of policy professionals. Policy professionals who see behaviour as more fixed were hypothesised to prefer strict policy tools (e.g., fines) over soft policy tools (e.g., nudges, provision of information) when aiming to change behaviour. The results show though that professionals seeing behaviour as more fixed prefer more strict *and* soft policy tools, thus generally more government effort to change behaviour. This chapter illustrates that the integration of behavioural insights into public policy also takes place in the minds of policy professionals.

Taking four years of experience with BIG'R as a basis, in **Chapter 8**, I speculate about different strategic dimensions for BITs how to integrate behavioural insights into public policy. These

dimensions concern the science-policy nexus, different forms of impact, and organisational formats of BITs.

## **Contributions**

I was responsible for the overall design of this research. This included the formulation of the research question and ensuring that the different parts of this research complemented each other in answering the research question. However, this research was completed using a collaborative working mode and benefited from contributions from different directions and perspectives, namely those of policy professionals working for the municipality and the academic scholars, especially my supervisors, who are listed as my co-authors.

I was the principal investigator for the research reported in the chapters that list me as the first author. This included taking responsibility for the research process, making design decisions, collecting and analysing data, and publishing the findings. I contributed to these tasks in a supportive role for the chapters that list me as the second author.





## Part 1

# Chapter 1

## Applying Behavioural Insights to Public Policy

### An Example From Rotterdam

This chapter has been published as:

Dewies, M., Denktaş, S., Giel, L., Noordzij, G., & Merkelbach, I. (2022). Applying behavioural insights to public policy: An example from Rotterdam. *Global Implementation Research and Applications*, 2(1), 53–66. <https://doi.org/10.1007/s43477-022-00036-5>

## Abstract

Worldwide, scholars and public institutions are embracing behavioural insights to improve public policy. Multiple frameworks exist to describe the integration of behavioural insights into policy, and behavioural insights teams (BITs) have specialised in this. Yet, it often remains unclear how these frameworks can be applied and by whom. Here, we describe and discuss a comprehensive framework that describes who does what and when to integrate behavioural insights into policy. The framework is informed by relevant literature, theorising, and experience with one BIT, the Behavioural Insights Group Rotterdam. We discuss how the framework helps to overcome some challenges associated with integrating behavioural insights into policy (an overreliance on randomised control trials, a limited understanding of context, threats to good scientific practice, and bounded rationality of individuals applying behavioural insights).

Worldwide, scholars and public institutions are acknowledging that a better understanding of human behaviour can improve public policy (e.g., Executive Order No. 13707, 2015; Wetenschappelijke Raad voor het Regeringsbeleid, 2014; World Bank, 2015). Drawing from research in the behavioural sciences, scholars and public institutions aim to understand why humans behave as they do and use these behavioural insights (BIs) to develop, test, and implement behavioural solutions for policy issues. Here, BIs refer to an understanding of behavioural determinants as well as how behaviour may be changed. The publication of the book *Nudge* (Thaler & Sunstein, 2008) can be seen as a renewed starting point and catalyst for the worldwide interest in the integration of BIs into public policy (Whitehead et al., 2018). This integration often includes the use of nudges: small and seemingly irrelevant changes in choice contexts that exploit behavioural automatisms such as biases, habits, and heuristics to programme behaviour (Thaler & Sunstein, 2008). To illustrate, BIs have been used to increase the rate of organ donors by changing from opt-in to opt-out defaults (Johnson & Goldstein, 2003), and to improve tax compliance by communicating social norms (“nine out of ten people pay their tax on time”; Hallsworth et al., 2017, p. 16).

Public institutions often employ so-called behavioural insights teams (BITs) as a centralised way of integrating BIs into public policy (as “public sector innovation labs”; McGann et al., 2018). These teams typically combine behavioural research and policy expertise to address policy issues with a behavioural dimension on a case-by-case basis (John & Stoker, 2019; Strassheim et al., 2015). BITs are often modelled after the first BIT which was set up within the United Kingdom Cabinet Office in 2010 (John, 2014). By now, several hundred BITs exist around the globe at various institutions (Afif et al., 2018; Manning et al., 2020; OECD, 2017). So far, little is known about the effectiveness of those teams, but evidence shows that interventions designed by BITs can result in effective policy changes (Benartzi et al., 2017; DellaVigna & Linos, 2020; Hummel & Maedche, 2019; Szaszi et al., 2018). Applying BIs in practice often goes hand in hand with a strong commitment to evidence-based practices of policy making and employing scientific methods to trial interventions (Einfeld, 2019; Haynes et al., 2012; John, 2014).

Multiple frameworks describe the integration of BIs into public policy. They have in common that they inform readers about *what* should be done for the successful integration of BIs and *when* it needs to be done. Illustrative and popular examples are the “easy, attractive, social, and timely” (EAST; Service et al., 2014) formula developed by the BIT in the United Kingdom, and the “behaviour, analysis, strategy, intervention, and change” (BASIC) step-by-step process developed by the Organisation for Economic Co-operation and Development (OECD, 2019). EAST answers the *what*-question by summarising the notion that to promote a target behaviour, interventions should make the behaviour easy to perform, appear attractive, be encouraged by the social environment, and stimulated in a timely manner. BASIC answers the *when*-question by prescribing a sequence of steps that lead to the integration of BIs. Namely, identification of a target behaviour, analysis of this behaviour, identification of behaviour change strategies, testing of interventions, and implementation of effective interventions to achieve change. Yet, these existing frameworks fall short to answer *who* needs to be involved in the integration, and the process of bringing it all together. Relatedly, there is

little research on how BITs operate and use those frameworks (Ball & Head, 2021). As a consequence, practitioners are often left wondering who should do what and when for the successful integration of BIs into policy. This paper aims to overcome this shortcoming by introducing a more comprehensive framework.

This framework relates to implementation science (e.g., Nilsen, 2015; Ogden & Fixsen, 2014) because it relies on evidence-based practices and their uptake in policy practice. Specifically, it can be interpreted as a determinant framework as well as a process model which outlines implementation steps. To the best of our knowledge, this is the first attempt to bridge implementation science and the emerging field of behavioural public policy.

We believe this study to be relevant for BIs practitioners, as well as scholars interested in behavioural public policy and behavioural public administration because the framework tackles some common challenges associated with the integration of BIs into policy that we summarise below.

## **BIs Challenges**

Based on the literature we identify four challenges associated with the integration of BIs into policy: (1) an overreliance on randomised control trials; (2) a limited understanding of context; (3) threats to good scientific practice; and (4) bounded rationality of professionals applying BIs. Strategies for overcoming the challenges are described when presenting the framework and in the Discussion.

### ***Overreliance on Randomised Control Trials***

From early on, the integration of BIs into policy within BITs has been coupled with the use of randomised control trials (RCTs) to build up knowledge about what works (Einfeld, 2019; Halpern & Mason, 2015; John, 2014). The coupling implied for some BITs that they would only conduct projects where a RCT was possible (Ball & Head, 2021). However, there is increasing recognition that integrating BIs into policy is not conditioned on RCTs and that the scope of integrating BIs increases once BITs also embrace other evaluative methods (Ball & Head, 2021; Einfeld, 2019; Ewert & Loer, 2021). Relatedly, an overreliance on RCTs has been associated with an insufficient appreciation of local and contextual factors affecting effectiveness of interventions (Pearce & Raman, 2014) which in turn limits the capacity of BITs to learn about working mechanisms and tailor interventions.

### ***Limited Understanding of Context***

Behaviour is embedded in and contingent on various levels of context reaching from the immediate material and social context to larger social structures, as well as historic and developmental dynamics (e.g., Bronfenbrenner, 1977). However, current practices of integrating BIs into policy tend to simplify or neglect contextual features that are more remote in space and time (e.g., early childhood, institutional contexts) for the analysis of behaviour, efforts to change it, and evaluations of interventions. Based on over 100 interviews with BIs experts, Whitehead and colleagues (2017; 2018) conclude that the neglect stems from an overreliance on knowledge and methods from behavioural economics instead of other relevant disciplines embracing broader understandings of context, such as

sociology and geography. In a similar vein, Muramatsu and Barbieri (2017) argue that interventions should not be based on overgeneralisations but be designed on a case-specific basis, integrating for instance knowledge from practice. In addition, multiple authors have highlighted the importance of diagnostic research to develop interventions that are tailored to their contexts (e.g., Hauser et al., 2018; Meder et al., 2018; Strassheim, 2019; Weaver, 2015).

### ***Threats to Good Scientific Practice***

Bolton and Newell (2017) have outlined five threats to good scientific practice for BITs. These threats stem from the culture of public institutions and include the following: (1) the pressure to develop effective solutions at low-cost; (2) the pressure to generate quick results and meeting deadlines; (3) inappropriate power assertions corrupting public policy or science; (4) difficulties to meet open science standards; and (5) limitations related to adequate operationalisations and good research designs. The last two threats have also been highlighted by the BIT in the United Kingdom (Sanders et al., 2018). In addition, several authors have highlighted the poor quality of reporting of trials executed by BITs (e.g., Cotterill et al., 2021; Osman et al., 2018).

### ***Bounded Rationality of BIs Professionals***

Integrating BIs into policy is often presented as an effort to stimulate rational behaviour that compensates for the bounded rationality of individuals (i.e., the notion that decision-making is impaired by limited cognitive capacities; Thaler & Sunstein, 2008). Yet, this has provoked scepticism concerning the capacity to steer rational behaviour when those steering it are not fully rational themselves (Lodge & Wegrich, 2016; Thomas, 2019). Indeed, individual personality traits, cue-taking, fast and intuitive thinking, and emotions were said to limit the potential for rational policy design (Nørgaard, 2018). In addition, the straightforward and rational design processes propagated by BITs were found to be adapted and subjected to organisational logics when introduced to policy practice (Feitsma, 2020) further feeding the scepticism concerning the capacity of BITs to act rational.

## **Comprehensive Integration Framework**

Our comprehensive framework for the integration of BIs into policy making and policy implementation aims to overcome the challenges. The framework describes the ingredients required for the integration and the procedure for combining these ingredients. Following the presentation of the framework, we illustrate how it was put to practice within one BIT, the Behavioural Insights Group Rotterdam.

### ***Development of the Framework***

The framework was developed in an iterative process of consulting relevant literature, theorising, and practical experience. Initially, the then available literature about BIs and BITs (e.g., Ames & Hiscox, 2016; Congdon & Shankar, 2015; Halpern, 2015; Halpern & Sanders, 2016; Haynes et al., 2012; Service et al., 2014; Thaler & Sunstein, 2008) was used eclectically to theorise a preliminary framework. Thaler and Sunstein (2008) can be seen as a groundwork highlighting the importance of

avoiding misconceptions about behaviour when formulating or implementing policy. To overcome such misconceptions, the idea of using a linear and rational procedure for integrating BIs into policy was based on the procedures of other BITs (Ames & Hiscox, 2016; Service, 2014). A focus on generating evidence and conducting trials was taken from other BITs as well (Congdon & Shankar, 2015; Haynes et al., 2012), however, it was complemented with a pragmatic perspective employing different evaluative methods (Morgan, 2007; 2014). The framework then guided more than three years of practical experience within one BIT, the Behavioural Insights Group Rotterdam (BIG'R), during which the framework was regularly discussed with BIG'R members and updated.

### ***Ingredients***

We differentiate between the ingredients behavioural research, policy expertise, and collaboration efforts. Competences needed for behavioural research and policy expertise are associated with different professions and individuals, making collaboration efforts to stimulate interaction and coordination between them a necessity. Integrating knowledge from behavioural research and policy expertise helps to overcome a limited understanding of context. The ingredients are displayed in Figure 1.1.

#### ***Behavioural Research***

To address policy issues using BIs, such issues need to be translated and reframed as behaviour change tasks (e.g., “too much energy is produced” to “individuals shower too long and hot”). This requires an understanding of what constitutes behaviour, the identification of target behaviours that underly policy issues, and behavioural terminology (Olejniczak et al., 2020). Inherently linked to this is the conduction of diagnostic research to better understand the target behaviour, its determinants, and its context utilising local knowledge. Bringing about behaviour change also requires knowledge about robust behaviour change techniques that likely work in the respective context and can guide the design of interventions. Some of these techniques are popularised in frameworks like EAST (Service et al., 2014). However, these frameworks rarely explain when which technique works, making expert behavioural knowledge a necessity to choose or design interventions on a case-by-case basis.

Similar to interventions, the evaluation of interventions needs to be adjusted to their context and take into account practical research constraints (e.g., limited resources, access to participants). As practical constraints can determine what kind of evaluation research is feasible, a pragmatic approach to research (Morgan, 2007; 2014) using multiple methods allows one to adjust the research to its context and overcome the overreliance on RCTs to evaluate interventions. In the context of BITs, research is often contingent on public servants for its support and assistance (e.g., entry to the field and assistance with data collection). Whatever research conducted, it should adhere to criteria for good scientific practice and research protocols of the public institution employing the BIT. Adherence to these criteria and protocols requires scientific training and expert knowledge (i.e., professional researchers) as well as independence of researchers, peer supervision, and quality control of the

conducted research. Therefore, BITs need to have an organisational mandate to enforce and ensure adherence to these criteria and protocols.

### *Policy Expertise*

Integrating BIs into policy requires expertise from policy for various reasons. Specifically, policy expertise is required to identify and select policy issues with a behavioural dimension of societal and administrative relevance. These issues need to be associated with distinct target behaviours to reframe the issues as a behavioural change task. Once issues have been selected, policy expertise (e.g., from front-line workers, policy advisors) helps to identify relevant stakeholders of issues to ensure their adequate involvement. Policy expertise and stakeholder input together help to better understand the context of target behaviours to tailor interventions, and take into account the local policy and administrative landscape when conducting research and testing interventions. Moreover, practitioners' knowledge often helps to avoid common mistakes and learn from earlier attempts to change the behaviour (Halpern & Sanders, 2016).

Ultimately, results from intervention evaluations need to feed into the design of actionable policy proposals. For this, behavioural research needs to be translated and embedded in the local administrative and policy landscape (Feitsma, 2019a). Only thereafter, can effective interventions be implemented using adequate policy tools to enable and maintain behaviour change (Michie et al., 2011). For this, policy expertise is required to secure an administrative mandate and support (e.g., budget) and advise throughout the implementation process. The former includes advocating for the integration of BIs in general and the implementation of policy proposals in particular, both using strategies of policy entrepreneurs (e.g., selling the effectiveness of behavioural interventions, identifying policy issues of political importance; Huijtema & Meijerink, 2010). To maximise the impact of policy proposals, policy expertise can identify opportunities for transfer as well as upscaling of proposals. Similarly to the behavioural research ingredient, policy expertise requires supervision and quality control of outputs (e.g., policy proposals).

### *Collaboration Efforts*

As a third ingredient, collaboration efforts are needed to combine behavioural research and policy expertise (Halpern & Sanders, 2016). This ingredient encompasses strategic direction, management, communicative skills and mutual understanding, mediation, and a standardised working method. Concerning the first, strategic direction establishes goals and how to achieve them. For instance, in its early days the BIT in the United Kingdom focussed on "quick wins" where minor changes promised to produce obvious revenues or savings to get political buy-in and support (Halpern, 2015; Sanders et al., 2018). Such a strategy ensures that different team members work towards a shared goal. Second, the team, its people, budget, and stakeholders need to be managed actively. This task combines the delivery of research projects with a focus on public management and requires "publicly engaged research managers", with important skills such as "diplomacy and negotiation" (Dunleavy et al., 2019).

Information does not flow between scientists and public servants in a linear and straightforward manner. It first needs to be translated and contextualised (Feitsma, 2019a). Consequently, as a third aspect of collaboration efforts, members of BITs are required to possess strong communication skills and a basic understanding of both science and policy. In case of disagreements and conflicts (e.g., concerning timeframes), it may even require mediation between researchers and policy experts, or a shared authority that can be consulted. At the BIT of the Australian Government for instance, there was a dual management with a Research Director and a Managing Director more embedded in the policy context as a shared authority (Ball et al., 2017). Involving both behavioural research and policy expertise assures that experiments are “co-produced” (Voß & Simons, 2018) aiming to solve local policy problems rather than translating abstract knowledge to fit generalised problem definitions (Gibbons et al., 1994). Finally, a standardised procedure or working method that instructs researchers and policy experts can ease collaboration. Such a procedure is outlined in the following section. It also helps to overcome the bounded rationality of BIs professionals because it provides a systematic approach to first explore a policy issue before proceeding to solutions.

### ***Procedure***

The procedure specifies how the ingredients from the framework are to be combined. It encompasses four separate phases that together form a logical and temporal sequence where the results from earlier phases feed into later ones (see Figure 1.1). To be more precise, it prescribes a sequence of (1) policy issue selection; (2) efforts to understand the underlying behaviour; (3) design of policy proposals, and (4) implementation and evaluation.

#### ***Policy Issue Selection***

The objective of the policy issue selection phase is to select adequate policy issues. As mentioned above, a BIT’s strategy may have a significant impact on which policy issues are selected. For the selection, a snapshot of the policy landscape related to a policy issue is needed to judge its relevance. Moreover, an initial reflection on the behaviour, its plausible determinants, and its context are required to judge the potential to effectively change the behaviour. In most cases, there will be multiple behaviours underlying a policy issue and BITs will need to focus on a small number of them given available resources. Furman (2016) has argued that starting with a policy issue (rather than starting with BIs and then looking for a suitable policy issue) holds the potential for BIs to play a more important role in public policy.

The use of the framework presumes the willingness of public servants to collaborate. The Australian BIT ensured this willingness by co-funding from project partners which “encouraged a stronger sense of engagement” (Ball et al., 2017). Ensuring buy-in from public servants and their support (e.g., financially) is thus of high importance especially since structural long-term partnerships, and equality and reciprocity between partners were found to ease implementation (e.g., Ansell et al., 2017; Torfing, 2019).



### *Exploration*

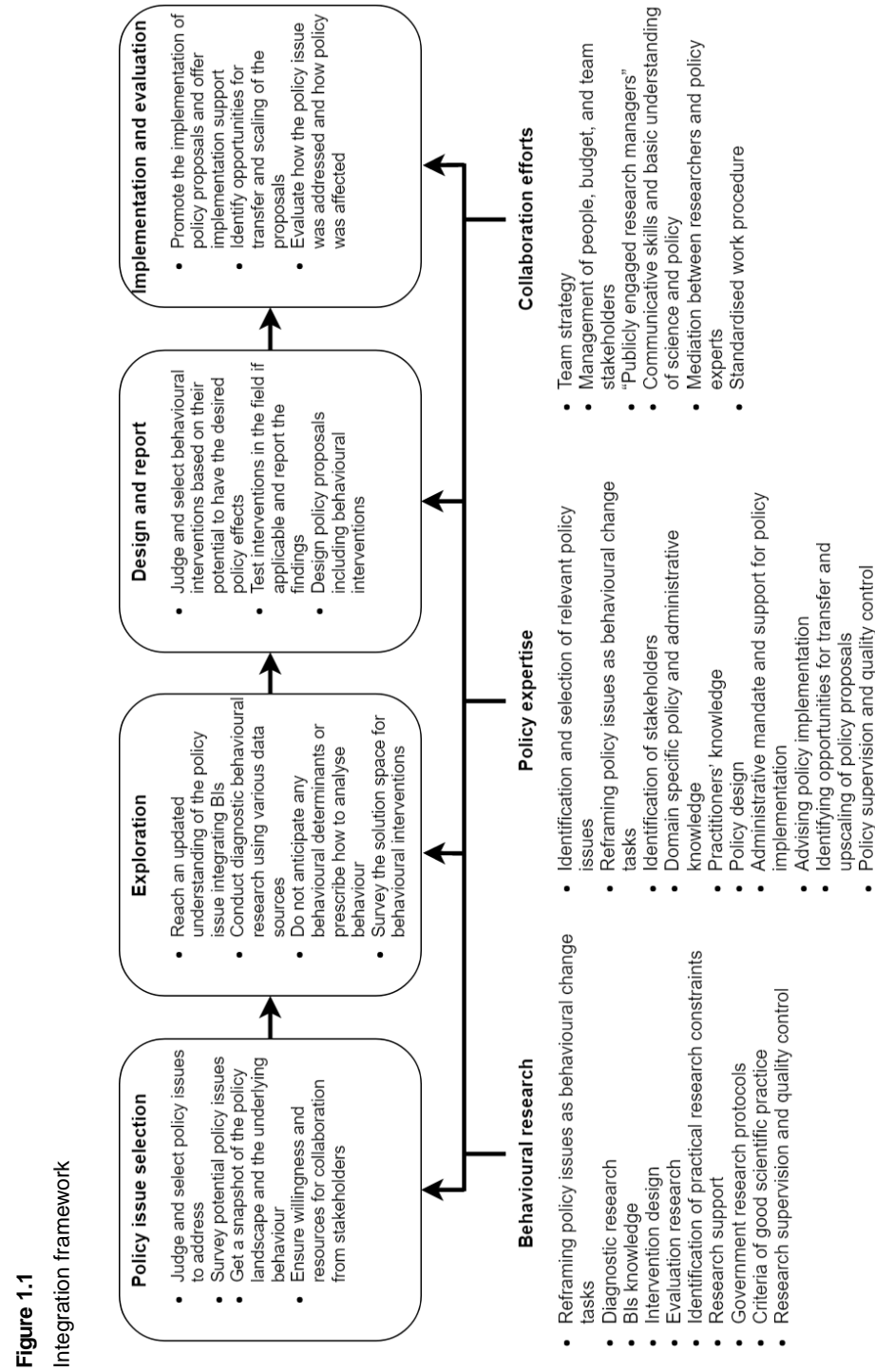
The objective of the exploration phase is to reach an updated understanding of the policy issue by integrating BIs into this understanding. Research has shown that public servants sometimes use wrongful assumptions when making inferences about behaviour (Hall et al., 2014; Howlett, 2018) and, in addition, that problem definitions based on singular perspectives increase the likelihood of solving wrong problems or non-problems (Kilman & Mitroff, 1979). During this phase, diagnostic behavioural research is hence employed making use of various data sources (e.g., observations, interviews) to better understand the target behaviour. To integrate findings from this diagnostic research with policy expertise, findings are validated collaboratively. This step helps to overcome the bounded rationality of BITs using collective scrutiny.

The diagnostic purpose of this phase overlaps largely with, for instance, the Diagnosis step from the guidelines of the Australian BIT (Ames & Hiscox, 2016) and the Analysis step from BASIC (OECD, 2019). In contrast to both, the framework does not anticipate any behavioural determinants or prescribe how to analyse behaviour. This way it can accommodate various behavioural determinants (e.g., choice related, sociological) opening up a large solution space that targets various determinants. To survey this solution space, BITs can use an analytical approach addressing relevant determinants or a designerly approach using participative and creative techniques to generate novel intervention ideas (for a discussion of both approaches see Einfeld & Blomkamp, forthcoming).

### *Design and Report*

The objective of the design and report phase is to design policy proposals that include behavioural interventions and to report research findings. Note, that the first objective is located in the policy sphere implying a responsibility of the BIT that goes beyond the delivery of behavioural interventions alone (Congdon & Shankar, 2015; Hansen, 2018). Rather, interventions need to be translated into policy proposals which are embedded in the institution's policy and administrative landscape. The literature on knowledge utilisation has identified abstractness as an important barrier for successful utilisation (Rich, 1997). As such, translating interventions to actionable policy proposals is thought to increase the likelihood that decision-makers adopt such proposals compared to a situation where technical evaluations of interventions would be shared (Mead, 2015).

To translate interventions into policy proposals, the best intervention ideas from the exploration phase must be selected. To address the policy issue, the selected interventions should have the potential to not only affect the target behaviour (i.e., statistical relevance) but also have practically meaningful effects on policy outcomes (i.e., clinical significance). In addition, BITs may use other relevant criteria such as affordability, practicability, cost-effectiveness, acceptance, side-effects, and equity (Michie et al., 2011).



With many BITs committed to a scientific agenda and evidence-based practices, they often trial some of their interventions in the field (Haynes et al., 2012). As a consequence, a differentiation can be made between policy proposals that are informed by such field trials (“behaviourally-tested”) and those that are informed by a review of pre-existing evidence (“behaviourally-informed”; Lourenço et al., 2016, pp. 15–16). For a proper reporting of trials, BITs can publish in peer-reviewed academic journals and/or use standardised reporting checklists (Cotterill et al., 2021). For this, it is important that publishing decisions and methodological choices are untouched by unqualified interests (e.g., political) and that conduction of the research is supervised by independent researchers.

### *Implementation and Evaluation*

The objective of the implementation and evaluation phase is twofold. The first objective is to achieve implementation of policy proposals produced during the design and report phase. Acknowledging that implementation is not a onetime event (e.g., sharing the policy proposal), BITs need to support and stimulate an implementation process (Ogden & Fixsen, 2014). For instance, they may need to compete for the attention of decision-makers or advice on adaptations of the proposal. As mentioned earlier, implementation can be improved by collaborating with stakeholders from early on and relying on the strategies of policy entrepreneurs. Moreover, BITs may actively look for transfer and scaling opportunities. It is beneficial for implementation when the same individuals continue to be involved during this phase, because they tend to have developed an understanding of the negotiables and non-negotiables for successful scaling (Al-Ubaydli et al., 2021).

The second objective of this phase is to evaluate how the policy issue had been addressed (i.e., a process evaluation) and how policy was affected (i.e., an outcome evaluation). Conducting such evaluations in a formative way (i.e., informing development) can help BITs to learn and adapt over time. This is important since many new BITs are likely to initially follow popular role models of integrating BIs into policy which, however, often need adjustments when introduced to practice (Feitsma, 2019a; 2020). Hence, evaluation and active learning can stimulate a process of diversification where BITs adapt to their contexts.

## **Illustration of the Framework**

As mentioned before, the framework originates in part from the authors’ experience with BIG’R. BIG’R was a collaboration between the Erasmus University Rotterdam and the municipality of Rotterdam with both institutions acting as equal partners (i.e., a “boundary organisation”; Guston, 2001). All authors were BIG’R members for at least one year as university researchers or the academic head. In the following, we illustrate how the framework was used within BIG’R. Specifically, we discuss how different job functions provided the ingredients for the framework and how the BIG’R working method relates to the four phases of the procedure for combining the ingredients. In addition, we refer the reader to the online supplementary material (<https://doi.org/10.1007/s43477-022-00036-5>) for a description of how the framework was applied to one policy issue.

### ***Ingredients: The BIG'R Job Functions***

The BIG'R job functions that were relevant for the framework were the BIG'R management, case proposers, municipality and university researchers, and policy domain advisors. To address policy issues on a case-by-case basis, representatives from these functions except the BIG'R management collaborated in so-called policy case teams. Each team consisted of at least two professionals contributing policy expertise (the case proposer and the policy domain advisor) and two researchers who provided checks and balances that could help neutralise the threats from inappropriate power assertions corrupting public policy or science.

#### ***BIG'R Management***

The BIG'R management consisted of the municipal project leader and the academic head. The former was an experienced project leader from the municipality and the latter a behavioural science professor experienced with field research. This way, the management was able to combine the technical and conceptual knowledge (Katz, 1955) from science and policy that was needed for the selection of policy issues, supervision of BIG'R members, and quality control. The BIG'R management also served as a shared authority for all BIG'R members. By managing the budget, interaction with stakeholders (e.g., municipality alderman), and operations, the management contributed collaboration efforts, especially since it made many decisions employing a participatory rather than a top-down approach. In practice this meant that daily operations were discussed and coordinated during a weekly meeting with representatives from all BIG'R functions.

The management also determined the strategy of BIG'R. Specifically, it set an emphasis on innovation and setting up BIG'R for the first two years, learning how to integrate BIs into policy; and a more production and implementation focussed working mode for the following two years. Importantly, the academic head as part of the management had the authority to create a proper research environment so that threats to good scientific practice could be overcome.

#### ***Case Proposer***

Any Rotterdam public servant could become a case proposer by submitting a policy issue with a behavioural dimension that was approved by the BIG'R management. To facilitate later implementation, BIG'R ensured at the start that proposers had the mandate and resources for addressing the policy issue (e.g., budget). The collaboration with BIG'R was free for case proposers, however. The case proposers sensing a problem and initiating the policy case ensured their willingness to collaborate.

Case proposers were important members of policy case teams attending meetings and contributing their experience and knowledge about the policy issue, including knowledge about relevant stakeholders. Typically, case proposers had several years of experience with their position and hence the policy issue. This way, the close collaboration with case proposers helped to overcome the challenge of a limited understanding of context. Indeed, from our experience regular interactions were a necessity for the successful completion of policy cases. If the policy case team conducted field

research, case proposers identified practical research constraints (e.g., sample size limitations) and provided research support (e.g., obtaining informed consent).

### *Researchers*

BIG'R employed university researchers as well as researchers seconded from the municipality's research department. All researchers had at least a university degree in a social and behavioural science discipline (e.g., psychology, public health, economics, pedagogy) and were trained in qualitative and quantitative methods. Both types of researchers collaborated on shared research tasks (i.e., diagnostic research, intervention design, intervention evaluation) providing checks and balances. Municipality researchers were in addition responsible for writing policy proposals and other tasks more in the interest of the municipality, particularly contributing knowledge from municipality research and ensuring adherence to municipality research protocols. University researchers were responsible for scientific tasks, particularly contributing scientific knowledge and expertise, meeting the criteria of good scientific practice, and publishing findings in peer-reviewed journals. In an attempt to overcome threats to good scientific practice, their professional responsibilities were with the university only. They were regular members of their department and subjected to the departmental hierarchy and regulations. However, as they regularly worked at and with the municipality, their research can be described as "embedded research" (Ward et al., 2021a; Ward et al., 2021b).

### *Policy Domain Advisor*

The municipality administration encompassed seven subdivisions focussing on relevant aspects of Dutch local governments (e.g., work and income, urban development). In acknowledging the scattered nature of knowledge (Muramatsu & Barbieri, 2017), BIG'R recruited one public servant from each of the seven subdivisions to become a part-time policy domain advisor (PDA) for BIG'R providing specialised policy expertise. At the same time, all PDAs continued with their work (e.g., as legal consultant) within their subdivisions to maintain their network and information channels. From there, PDAs recruited case proposers and identified relevant policy issues that could be addressed by BIG'R.

PDAs stood in-between case proposers and researchers, mediating between them to integrate needs from research and policy. They can be described as knowledge brokers (Feitsma, 2019a) with a strong focus on bridging, translating, and facilitating. This way, they helped contextualising both the research and policy proposals in the current policy and administrative landscape whilst also being responsible for writing policy proposals. The focus of PDAs was more process-oriented and to ensure the progress of policy cases. This required project management competencies (e.g., coordination and planning). In addition, PDAs identified practical research constraints and provided research support when needed.

PDAs can be described as policy entrepreneurs aiming to achieve policy change within their policy domains. Specifically, they advocated the implementation of BIG'R policy proposals and the application of BIs in general. For instance, they gave presentations at municipality fairs or informed the senior management about BIG'R. Importantly, describing PDAs as policy entrepreneurs adds the

notion of a promotional self-interest to the rather neutral image of the knowledge broker. We believe this aspect to be important for a BIT to achieve change (for an illustration see John, 2014).

### ***Procedure: The BIG'R Working Method***

The BIG'R working method equipped the four phases of the framework with distinct steps to be carried out and embedded the framework in the administrative context of the municipality.

After a Rotterdam public servant had proposed a policy issue, a BIG'R researcher and the PDA, whose subdivision the issue concerned, interviewed the case proposer and conducted desk research to fill in a submission form (see online supplementary material at <https://doi.org/10.1007/s43477-022-00036-5>) that provided a snapshot of the behaviour, its plausible determinants, its context, and its relevance. The form served as input for the BIG'R management to select policy cases based on their societal, administrative, and scientific relevance. The latter aspect was important because BIG'R aimed to publish its findings in academic journals.

Starting with a policy issue rather than a specific BI allowed BIG'R to address a diverse set of policy issues from various stages of the policy cycle (e.g., ex-ante policy appraisal, policy design, policy implementation) and with differing scopes (e.g., one-off behaviours, culture change, intergroup behaviours). Some of them were underexplored by behavioural sciences, such as individuals disposing fat and oil into the sewage system causing expensive blockings (e.g., Mattsson et al., 2014). Yet, requiring case proposers as a starting point meant that most policy issues related to the micro level from proposers' individual responsibilities (e.g., improving a specific letter) rather than the macro level (e.g., improving general municipal communication).

At the start of the exploration phase, what was known about the target behaviour was largely based on the case proposer as an informant. Diagnostic research was therefore often necessary to reach an updated understanding of the policy issue that integrated BIs. For this, BIG'R reviewed the related literature and often conducted field observations and/or interviewed various stakeholders. To summarise the updated understanding, the policy case teams used different visualisations and descriptions on a case-by-case basis (e.g., user journeys and fuzzy cognitive maps; Kosko, 1986). The fact that BIG'R identified cultural and sociological determinants for multiple cases illustrates that anticipating only choice related determinants can limit the understanding of target behaviours. For instance, when investigating the illegal use of fireworks in a football stadium, BIG'R identified cultural norms and values as major determinants that could hardly be changed using a behaviour change approach (chapter 5 of this dissertation).

The face validity of this understanding and the prospects of a behavioural solution were in most cases then validated during a brainstorming session with other BIG'R members and relevant stakeholders. For one case, the understanding was judged insufficient to proceed during such a session, illustrating that this step helped to overcome the bounded rationality of BIs professionals. In addition, brainstorming sessions served to generate novel intervention ideas and to improve interventions known from the literature or practice. To stimulate creativity during the sessions, participants were primed (Rietzschel et al., 2007) using common BIs. In practice this could mean that

they were asked to think of solutions making the target behaviour easier (the “e” in EAST; Service et al., 2014).

Policy case teams selected interventions to be included in policy proposals afterwards based on their feasibility (e.g., financial, acceptance) as well as believed effectiveness. Although BIG’R did not pre-commit to any type of interventions, in practice the policy proposals often included nudges (Thaler & Sunstein, 2008). Beyond nudges, BIG’R for instance suggested a participatory programme to define a code of conduct for the use of electric charging stations (see online supplementary material at <https://doi.org/10.1007/s43477-022-00036-5>). Whenever possible, BIG’R tested (some aspects of) the proposed interventions in the field using both quantitative and qualitative methods (e.g., chapter 3 of this dissertation).

Policy proposals were presented and discussed with case proposers and other relevant municipality decision-makers to address open questions and ease implementation. However, it was a learning process for BIG’R to prepare policy proposals rather than technical research reports. In the end, proposals sometimes took the format of infographic-like summaries because from our experience case proposers demonstrated little interest in technical aspects of the research, plausibly because most case proposers had not received science training to understand and scrutinise the research. This meant that in practice, case proposers relied on authority-based rather than evidence-based policy making (Mendel, 2018).

In addition to policy proposals, the university researchers aimed to report findings in peer-reviewed academic journals. Importantly, the university and the municipality confirmed that methodological choices and publishing decisions would be the responsibility of the university to overcome threats to good scientific practice. The municipality and the university also agreed on a formal framework for sharing data between the two institutions that included the right to publish anonymous data.

It was beneficial for the implementation of policy proposals that BIG’R was embedded in the municipality administration, and that PDAs could approach proposers naturally after the proposal had been shared. Moreover, PDAs held a legitimate position for advocating the use of BIG’R proposals within their subdivision. PDAs hence acted as policy entrepreneurs during this phase. Ideally, they became part of “implementation teams” and identified opportunities for upscaling. Researchers conducting the original research could easily be consulted for this by the PDAs if needed.

To evaluate the BIG’R working method, the process as well as the outcomes were systematically evaluated on a case-by-case basis (see chapter 6 of this dissertation). Specifically, policy case team members were invited to fill in questionnaires covering process aspects of the collaboration (e.g., satisfaction with the collaboration and the policy proposal) and case proposers were interviewed to investigate the degree of implementation of policy proposals. However, this evaluation was more summative (i.e., informing assessment) than formative as it took place during the last year of BIG’R.

**Table 1.1**

Challenges and solutions for integrating BIs into public policy

<i>Challenges</i>	<i>Solutions</i>	<i>Strengths</i>	<i>Weaknesses</i>
<i>Overreliance on RCTs limits the scope of applying BIs</i>	<ul style="list-style-type: none"> <li>– Pragmatic approach to research with no a priori choice of research methods</li> <li>– BITs select research methods on a case-by-case basis</li> </ul>	<ul style="list-style-type: none"> <li>– Broadening the scope for applying BIs</li> <li>– Enabling more realistic evaluations of interventions</li> </ul>	<ul style="list-style-type: none"> <li>– Increased efforts to construct and make methodological choices</li> <li>– Researchers are required to be methodological generalists</li> </ul>
<i>A limited understanding of context limits effective intervention design, tailoring, and evaluation</i>	<ul style="list-style-type: none"> <li>– Standard involvement of policy experts and practitioners</li> <li>– Diverse disciplinary backgrounds</li> <li>– No a priori framework for interpreting behaviour and changing it</li> </ul>	<ul style="list-style-type: none"> <li>– Contextualised intervention development and evaluation</li> <li>– Production of contextualised policy proposals</li> <li>– Expanding the range of interventions</li> </ul>	<ul style="list-style-type: none"> <li>– Increased efforts to investigate theories, research findings, and behaviour</li> </ul>
<i>Threats to good scientific practice are likely within BITs</i>	<ul style="list-style-type: none"> <li>– Researchers as part of the management</li> <li>– Independence of researchers</li> <li>– Formal agreement about data sharing practices, methodological choices, and publishing decisions</li> <li>– Working in teams to provide checks and balances</li> </ul>	<ul style="list-style-type: none"> <li>– Integrity of researchers remains intact</li> <li>– Open science practices are enabled</li> </ul>	<ul style="list-style-type: none"> <li>– Increased bureaucracy and obligations</li> </ul>
<i>Bounded rationality of BIs professionals limit the potential to stimulate rational behaviour</i>	<ul style="list-style-type: none"> <li>– Working in teams and with stakeholders to provide checks and balances</li> <li>– BIs as an additional perspective</li> <li>– A standard procedure for policy issues</li> </ul>	<ul style="list-style-type: none"> <li>– Enhanced problem-solving capacity</li> </ul>	<ul style="list-style-type: none"> <li>– Requires extra investments</li> <li>– Increased potential for human error and biases</li> </ul>



## Discussion

This paper offers a framework for integrating BIs into the making and implementation of public policy. The framework is depicted in Figure 1.1 and describes the what, when, and who of integrating BIs into public policy. We believe that the framework is more complete than earlier frameworks which often lack descriptions of actors in the integration. Moreover, the framework incorporates ways to overcome four common challenges associated with the integration of BIs that we discuss below (for a summary see Table 1.1).

However, as the framework is closely related to our experience and practice within BIG'R, we note that the framework is a solution but not necessarily *the* solution to overcome these challenges. Future applications of the framework likely require some adaptations. As an example, BIG'R differed from other BITs because it was run by both government and academia. Most other BITs are run by either of these institutions. For instance, BITs run by government likely choose different projects because they can disregard their scientific relevance, and they likely focus more on using BIs rather than generating novel BIs by trialling them in the field. We hence consider the university partnership key for important aspects of the framework.

### ***Overcoming an Overreliance on RCTs***

The framework allows BITs to embrace various research methods. It does not deterministically lead to the planning and execution of RCTs, but encourages users to reflect on their methodological choices on a case-by-case-basis. Ewert (2020), and Ewert and Loer (2021) have recently argued for such a methodological diversification in the field of behavioural public policy. In fact, BIG'R conducted multiple cases where RCTs were not feasible or inadequate. This strength came, however, at a cost. First, it required increased efforts in terms of time and interaction for policy case team members to explore and choose research options. In addition, the framework required the researchers to be methodological generalists rather than RCTs specialists. We believe, however, that compared to a situation when only RCTs are used, opening a debate about the most adequate research method(s) leads to more informed and better methodological choices.

### ***Overcoming a Limited Understanding of Context***

From our experience with BIG'R, several features of the framework can help to overcome limited understandings of context, enabling the contextualised development and testing of interventions and writing of policy proposals. Importantly, the framework does not entail any (disciplinary) assumptions regarding how to interpret behaviour and how to change it. Earlier frameworks tend to anticipate specific behavioural determinants, namely choice related determinants on the micro level (e.g., food ordering), whilst excluding others, namely structural ones on the macro level (e.g., presence and dispersion of fast food restaurants). This neglect also means pre-committing to interventions that so far have dominated the integration of BIs into policy. In contrast to that, the framework presented here encourages the integration of divergent perspectives and sorts of knowledge from different disciplines. To illustrate, within BIG'R the case proposer typically contributed practitioner's knowledge, the PDA

typically contributed local policy knowledge, the municipal researcher contributed findings from local research, and the university researchers mainly contributed the BIs knowledge and expertise. In practice, it required, however, significant efforts to survey and integrate theories, research findings, and various perspectives on a case-by-case basis.

### ***Overcoming Threats to Good Scientific Practice***

Our framework contains multiple aspects that may help to overcome threats to good scientific practice. First, to strengthen the integrity and position of researchers, the framework assumes independence of researchers and an organisational mandate to enforce adherence to criteria of good scientific practice. Within BIG'R, the management possessed the mandate to create an adequate research environment as well as the knowledge that was required for that. Independence further means that methodological choices and publishing decisions for academic articles remain with researchers. Within BIG'R, professional liabilities of university researchers were with the university only which strengthened their position in relation to the municipality and policy practice. A disadvantage of being rigorous about scientific practices may be seen in the fact that all research becomes subject to scientific obligations. Within BIG'R for instance, requirements from the university for informed consents were often stricter than those from the municipality, and all empirical research had to be approved by the university's ethics committee which sometimes caused short delays in the execution of policy cases. The fact that science slowed down some policy cases can be interpreted, however, as a confirmation that science was not subordinated to municipal deadlines. Some obligations were, however, directly linked to positive outcomes. For instance, the obligations associated with scientific reporting and academic peer-review plausibly improved the reporting and evaluation of trials.

Formal agreements about data sharing practices enable open science practices (e.g., publication of anonymised data sets) to become a routine. Yet, for BIG'R a considerable amount of effort was needed to draft those agreements even with the help of legal experts from both institutions. Moreover, following the agreements caused some extra bureaucracy because forms needed to be filled in and approved by the management for each data sharing event. Some decisions related to good scientific practice still needed to be made on a case-by-case basis though, for instance decisions related to experimental designs.

### ***Overcoming Bounded Rationality***

With the lack of both a standard research methodology and no a priori assumptions about how to interpret and change behaviour, the framework deliberately shifts responsibility to team level decision-making. As a result, it can be said to increase the potential for human error and biases, and the risks related to bounded rationality (i.e., decision-making being impaired by limited cognitive capacities). In practice, however, following the framework means that policy issues are investigated from a behavioural perspective *as an addition* to the standard approach. Rather than fully adopting the policy

issue and addressing it with an exclusive focus on BIs, BIs are integrated with existing knowledge and experience concerning the policy issue. For BIG'R this meant that the administrative responsibility for the policy issue remained with the case proposer. Rather than the proposer addressing the policy issue in a standard approach, the BIG'R resources became available to scrutinise that standard approach, aiming for an approach that integrated BIs. We hypothesise that this enhanced the problem-solving capacity of the municipality locally (Gray, 2000; Rogers & Weber, 2010), and that the risks associated with rationally bounded public servants were lowered. However, this required making use of the resources invested in BIG'R.

### ***Comparison With Implementation Science Frameworks***

As the framework serves the implementation of evidence-based practices that are informed by BIs, the framework becomes available for a comparison with established and more general implementation frameworks (i.e., Damschroder et al., 2009; Fixsen et al., 2005). Admittedly, the framework lacks specificity and scope in comparison to those frameworks. Nevertheless, we believe the framework to be a valuable contribution because BITs possess a unique position at the science-policy nexus (Mukherjee & Giest, 2020; Strassheim, 2019) which plausibly requires them to utilise specialised frameworks. Borrowing terminology from the Active Implementation Frameworks (Fixsen et al., 2005), at least within BIG'R, "structural features for implementation" (i.e., creating dedicated implementation teams and developing implementation plans) were almost absent and dependent on processes stimulated by BIG'R or case proposers. A comparison with the Consolidated Framework for Implementation Research (Damschroder et al., 2009) points moreover to a lack of dedicated efforts to attract and involve appropriate individuals for implementation (e.g., opinion leaders, champions). This was different for the BIT in the United Kingdom for instance, which was located at the Cabinet Office and in comparison to BIG'R could more easily capitalise on its political support to achieve implementation (John, 2014; Lourenço et al., 2016). BIG'R deliberately aimed to stay away from political dynamics or affiliations to ensure its continued existence. However, this may have weakened its position for achieving implementation. Future use of the framework should take these aspects into account by for instance focussing on politically relevant policy issues.

### ***Conclusion***

The popular standard models of applying BIs propagated through the book *Nudge* (Thaler & Sunstein, 2008) and the BIT in the United Kingdom do not consider all relevant aspects of integrating BIs into public policy. Therefore, we developed a comprehensive framework that deviates from current models to overcome important challenges associated with integrating BIs into public policy, contributing to a recent phase of diversification in BIs integration (Strassheim, 2021). The framework offers more freedom for methodological choices, encourages and strengthens good scientific practice, incorporates a broad understanding of context, and improves the potential for rational decision-making.



## Part 2

# Chapter 2

## Call First, Pay Later

**Stimulating Debtors to Contact Their Creditors Improves Debt Collection  
in the Context of Financial Scarcity**

This chapter has been published as:

Dewies, M., Schop-Etman, A., Merkelbach, I., Rohde, K. I. M., & Denktas, S. (forthcoming). Call first, pay later: Stimulating debtors to contact their creditors improves debt collection in the context of financial scarcity. *Behavioural Public Policy*.

## Abstract

Debtors were stimulated to contact their creditors to negotiate a repayment plan. Contacting creditors was important because debtors were unlikely to repay the debt immediately and upon contacting, debtors could agree on a repayment plan to repay the debt in the long run. Using insights from scarcity theory and nudging techniques, a standard debt repayment letter was adapted and both letters were compared. Experimental results ( $N = 3,330$ ) provide support for the use of nudging techniques as more debtors agreed on a repayment plan and response rates increased. The results underline the importance of stimulating debtors to contact their creditors.

Recent insights from the behavioural sciences have highlighted the importance of contextual rather than personal factors for producing and maintaining financial hardship (Curchin, 2016; Gennetian & Shafir, 2015; Tiemeijer, 2016). For instance, application forms for financial support were found to be unnecessarily complex thereby reducing application rates and limiting opportunities for individuals to improve their situation (Bettinger et al., 2012). From this perspective, financial hardship is not just the result of individual deficiencies but also caused by social practices (Walker, 2012; Walker et al., 2015). This motivated public servants from Rotterdam in the Netherlands to look for improvements in a municipal repayment letter that informed some Rotterdam residents about a debt and asked them to repay.

The letter concerned the repayment of unjustified welfare payments the debtors had received from the municipality, for instance when the municipality was informed too late about additional incomes so that welfare payments had not been reduced or stopped in time. As a consequence of having received welfare payments until shortly before the letter was sent, the debtors were assumed to experience financial scarcity. In fact, pre-experimental data showed that most debtors did not repay debts higher than 100 euro immediately and in full, plausibly because of liquidity constraints. Here, we report the field testing of the improved repayment letter.

### ***Rationale for Improving the Repayment Letter***

In assuming that most debtors were bound by liquidity constraints (89% of debts were higher than 100 euro), the repayment letter was improved to increase the proportion of debtors that contacted the municipality after receiving the letter. Contacting was the first step towards agreeing with the municipality to repay the debt in the long run using monthly instalments. Contacting and agreeing on repayment by instalments also enabled debtors with other debt accounts of higher interest rates to prioritise repaying higher interest debts (Greenberg & Hershfield, 2019), thereby reducing overall expenses. Hence, contacting enabled the municipality to receive the repayment in a socially responsible and non-intrusive way.

Insights from scarcity theory (Mullainathan & Shafir, 2013; Shah et al., 2012) informed improving the repayment letter. Scarcity theory describes the mental consequences of having too little of important resources, in this case financial resources. According to the theory, scarce monetary resources lead to greater mental focus on monetary aspects (Shah et al., 2012; Shah et al., 2019). Having to pay back the debt is then more than just another task to complete; it is inducing complex financial trade-offs that demand extensive thought (Shah et al., 2015). As a consequence, fewer cognitive resources are available for other dimensions (Zhao & Tumm, 2017), i.e., to plan and conduct the repayment. Although there is convincing evidence for scarcity theory (e.g., de Bruijn & Antonides, forthcoming), some studies reported difficulties with replicating this evidence (Camerer et al., 2018; Shah et al., 2019). Nevertheless, we concluded that the cognitive demands originating from financial aspects and financial trade-offs implied in the letter should be minimised.

It can be illustrated that reducing cognitive demands extends beyond using more accessible language: Individuals experiencing financial scarcity are more likely to use their financial situation to

categorise themselves as poor, drawing an in-group and out-group distinction between themselves and their creditors (Akerlof & Kranton, 2000). This categorisation in turn increases cognitive demands to engage and interact with creditors (Mewse et al., 2010; Vivian & Berkovitz, 1992) as perceived out-group membership negatively influences the willingness to cooperate (Tajfel & Turner, 1979; Turner et al., 1987). Consequently, the letter could be improved by dissolving differences between debtors and creditors.

### ***Adaptations***

To minimize in- and out-group thinking, one of the adaptations to the letter was to stress togetherness and collaboration between debtors and creditors for finding a solution. An emphasis on collaboration can also lead to higher perceptions of procedural and interpersonal fairness which both increase motivations to cooperate (De Cremer et al., 2005; De Cremer & Tyler, 2007).

Two additional adaptations served to improve opportunities to contact employees of the municipality. First, an email-option was added for contact at any time of the day and with more anonymity (McFarland & Ployhart, 2015) and second, the generic telephone number of the municipality was replaced with the direct number of the responsible department. Debtors were assumed to expect better and more immediate help when talking to the departmental experts. In addition, action-related information was added so that debtors knew when they could call and what information they needed at hand. Action-relevant information can reduce uncertainty and increase confidence in being able to perform the respective action improving the translation into actual behaviour (Armitage & Conner, 2001).

Three final adaptations aimed to improve readability of the letter. First, choice options were presented as bulleted lists rather than as running text to improve memorising (Jansen, 2015). Second, pictograms were added to improve comprehension of the letter (Anglin et al., 2004; Levie & Lentz, 1982) and understanding of the letter's topic structure (Lorch, 1989; Lorch et al., 1993). Third, a concrete deadline (e.g., "10 June") was added to the relative deadline ("in six weeks") to free debtors from the need to pause reading for calculating the deadline or memorising to calculate it later, thereby allowing a more fluent reading experience. Task instructions that read more fluently were found to be associated with higher levels of estimated task easiness (Song & Schwarz, 2008) while lower levels of reading fluency have been associated with choice deferral (Novemsky et al., 2007).

These adaptations rely on nudging techniques: small and seemingly irrelevant changes in how choices are presented to exploit behavioural automatisms such as biases, habits, and heuristics to programme behaviour (Thaler & Sunstein, 2008). Similar nudging techniques have been used before to improve tax payments (e.g., Hallsworth et al., 2017) and reminders for debt repayment (Janssen et al., 2017; Jensen et al., 2018; Kondratjeva et al., 2021).

### ***Relevance***

The improved repayment letter was assumed to help some Rotterdam residents to better deal with indebtedness. Indebtedness is often perceived as stressful (Brown et al., 2005). Moreover, high levels



of indebtedness can have negative effects on life-satisfaction (Ruberton et al., 2016), well-being (Tay et al., 2017), and mental as well as physical health (Clayton et al., 2015; Fitch et al., 2011; Sweet et al., 2013). For the Netherlands, the percentage of households with low income, and consequently at risk of problematic indebtedness, was highest in Rotterdam (15%; Centraal Bureau voor de Statistiek, 2018). Adaptations in the letter promised to be cost-effective and easily scalable (Benartzi et al., 2017). Contacting behaviour of debtors has rarely been studied (for an exception see Mewse et al., 2010) and to the best of our knowledge this is the first study that stimulated contacting to ultimately improve debt repayment.

### ***Hypotheses***

We hypothesized that an improved repayment letter would stimulate more debtors to agree on a repayment plan. To investigate possible compensatory effects where more individuals agree on a repayment plan, but fewer individuals immediately repay the full debt, we analysed overall response rates as a secondary outcome measure. We hypothesized that overall response rates would increase with the nudging letter. In addition, we hypothesized that debtors agreeing on a repayment plan would adhere to the plan. Put differently, we assumed an indirect effect where more debtors agree on a repayment plan leading to an increase in repayments.

## **Methods**

### ***Design***




Given the municipal technical infrastructure, it was not possible to randomise the type of repayment letter at the individual level. Therefore, a quasi-experimental design was employed for this field research where the type of letter was dependent on the year it was sent. Specifically, we compared data from a period where the standard letter was sent (19 June 2017 till 8 November 2017) with data from the same period one year later where the nudging letter was used.

### ***Letters***

The standard letter and the nudging letter differed concerning the instructions for repayment. The remainder of the letter was the same in both letters and included information on why the debt was accrued and its amount. The two differing paragraphs are shown in Figure 2.1 and typically appeared in the centre of the first page of the letter. Normally, the letter was between one and a half and two pages long. The adaptations made in the nudging letter were the result of an internal brainstorm of the Behavioural Insights Group Rotterdam ([www.bigrotterdam.nl](http://www.bigrotterdam.nl)), a pilot test, and team discussion.

**Figure 2.1**

Translated excerpts from the standard and the nudging letter

	standard letter
<p><b>How do you transfer the amount?</b></p> <p>You have 6 weeks time to pay us back the amount. For this you receive a cheque. You can transfer the amount directly without a cheque to <i>[bank account number]</i>, Municipality Rotterdam Work and Income. Indicate then <i>[client's name]</i> and your social security number <i>[client's social security number]</i>.</p> <p><b>Can't you pay right now?</b></p> <p>You can arrange payment by instalment with us. Contact us as soon as possible for this: 14010.</p>	
<p style="text-align: right;"><b>nudging letter</b></p> <p><b>How do you transfer the amount?</b></p> <p>You have 6 weeks' time to pay us back the amount, thus till <i>[date]</i>. There are 2 modes of payment:</p> <div style="display: flex; align-items: flex-start; margin-top: 10px;"> <div style="margin-right: 20px; text-align: center;">  </div> <ul style="list-style-type: none"> <li>With a cheque that we send you later</li> <li>By transferring the amount yourself via internet banking to <i>[IBAN]</i>, Municipality Rotterdam Work and Income. Indicate then <i>[client's name]</i> and your social security number <i>[client's social security number]</i>.</li> </ul> </div> <p><b>Can't you pay right now?</b></p> <p>You also can arrange payment by instalment with us. Contact us as soon as possible so we can find a solution with you together:</p> <div style="margin-top: 20px;"> <div style="display: flex; align-items: center; margin-bottom: 10px;"> <div style="margin-right: 20px; text-align: center;">  </div> <p>Call between 9.00 and 12.00h <i>[department phone number]</i> department reclamation and indemnity for payment by instalments. Keep this letter at hand so we can help you more quickly.</p> </div> <div style="display: flex; align-items: center;"> <div style="margin-right: 20px; text-align: center;">  </div> <p>Or email to <i>[department email]</i>. State your administrative number in the subject line of the email: <i>[client's administrative number]</i>.</p> </div> </div>	

**Procedure**

The municipality of Rotterdam worked with a standard collection procedure for all debts resulting from work and income related services. Sending out the repayment letter marked the beginning of this procedure. After that, debtors were given 1.5 months (i.e., six weeks) to respond, either by paying back the whole debt or by contacting the municipality for negotiating a repayment plan. During that period, reminder letters were sent to some debtors initiated by the employees in charge of debt collection.

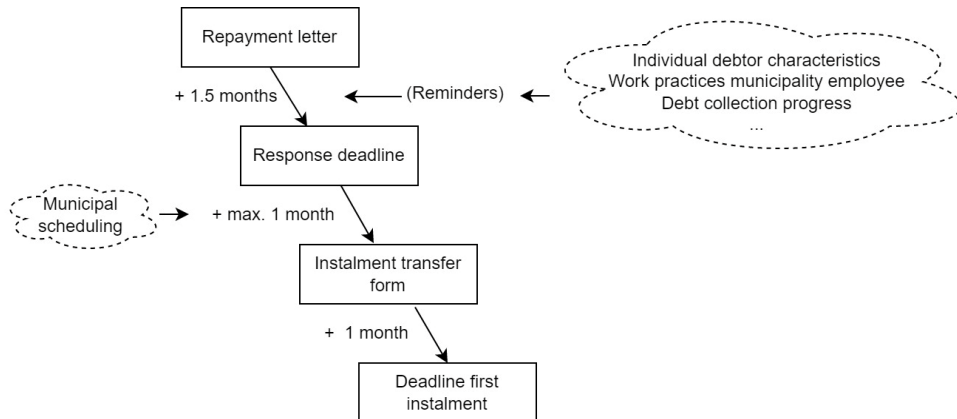
Whenever an agreement on a repayment plan had been reached, the municipality sent a letter of confirmation immediately and a money transfer form within one month, because those forms were only sent on one specific date for each month. The transfer form granted individuals one month's time for

their first instalment. If debtors agreed on monthly instalments, they were thus expected to transfer the first instalment within a maximum of 1.5 months plus two months after having received the repayment letter (Figure 2.2). If debtors had not responded after the response deadline, they were automatically sent multiple overdue notices and ultimately their wages could be garnished.

Data on all debtors and correspondence was registered and stored as part of the standard collection procedure of the municipality. The municipality of Rotterdam agreed to share datasets relevant for the evaluation of the nudging letter with the authors. Specifically, we received three datasets: one containing information on all debtors receiving the repayment letter, one containing metadata on correspondence other than the repayment letter, and one containing information on repayments. The data was cross validated by the first author for 50 randomly chosen debtors with the individual dossiers of the debtors at the municipality, in which all personal information and correspondence was stored. This validation revealed only two inconsistencies: one case where a reminder letter was wrongly used to confirm contact that had been made before and one case where a repayment plan was confirmed using a wrong letter type, meaning that this agreement was overlooked. The municipality had no role in the analysis and interpretation of data, in the writing of the article, and in the decision to submit it for publication. Approval to conduct this research was granted by the local ethics committee.

**Figure 2.2**

Schematic overview of the repayment procedure for repayment via instalments



## Measures

### *Agreeing on a Repayment Plan*

To measure whether or not debtors had agreed on a repayment plan, we constructed a variable that indicated whether debtors had received the letter confirming a repayment plan no later than 45 days after the repayment letter. The 45 days limit was chosen because it corresponded to the response deadline (i.e., 1.5 months) upon which debtors who could not pay were expected to contact the municipality plus an additional three days processing period before sending the confirmation letter.

### *Response*

The debtors who neither agreed on a repayment plan nor had made any repayment within 45 days after the repayment letter had been sent, were assumed to not have shown any response to the letter. We therefore constructed a response variable, which we set to one in case debtors had either agreed on a repayment plan or made any payment within 45 days after the repayment letter

### *Repayment*

We constructed a repayment variable to reflect repayment of (part of) the debt before the instalment deadline (see Figure 2.2). For each debtor it was checked if they had made at least one repayment not later than 105 days after the repayment letter had been sent. This repayment could reflect full repayment of the debt or instalments. The number of days corresponded with the maximum of 45 days for agreeing on a repayment plan plus maximum one month (i.e., approximately 30 days) for receiving the transfer form, and one month before the payment deadline mentioned on the transfer form.

### ***Analytic Strategy***

All analyses were carried out within the R statistical computing environment (version 4.0.2). The effect of the nudging letter on response was examined by a logistic regression. Specifically, we separately regressed the binary agreement on a repayment variable, and the response variable on the letter variable. To evaluate the model, we compared it against the null model using likelihood-ratio testing and inspected Nagelkerke's  $R^2$  (Nagelkerke, 1991). Moreover, we conducted sensitivity analyses using age and gender as covariates. Multicollinearity for these sensitivity analyses was assessed by inspecting the correlation matrix of the predictors and the variance inflation factors (VIF). We assumed multicollinearity to be a concern if correlations between predictors exceeded .80 and the VIF exceeded 2.5 (Midi et al., 2010). In another sensitivity analyses that was suggested by a reviewer, we investigated abrupt increases in the proportion of debtors agreeing on a repayment plan or responding to the letter around the date that the letter was changed (21 December 2017) using a regression discontinuity design and additional data sets.

Some debtors received a reminder from the municipality before the deadline to respond. We did not include this reminder as a covariate as it was more likely to be the result of the unresponsive behaviour of debtors rather than a determinant of a response. Moreover, compared to the debtors who received the nudging letter, more debtors receiving the standard letter received a reminder (see Table 2.1). Thus, any positive effect of the nudging letter when compared to the standard letter, could not be attributed to the effectiveness of reminder letters. Not including those letters as a covariate was therefore assumed to lead to a conservative estimate of the effect of the nudging letter.

To test the mediation effect of the nudging letter on repayment via contacting, we followed the procedure suggested by Zhao and colleagues (2010). They argued that to conclude mediation, only the indirect effect  $d = a \times b$  (see Figure 2.6) needs to be significant and that a significant direct effect  $c$  of the predictor on the outcome is not necessary. For this analysis, we employed the lavaan package

(version 0.6-5; Rosseel, 2012). Specifically, we used bootstrap testing drawing 2,000 samples in combination with structural equation modelling to test for the indirect effect. In this modelling structure, we conducted again a sensitivity analysis with the same covariates as with the logistic regressions.

In addition, we conducted an exploratory analysis in which we analysed whether debtors who had received the nudging letter had agreed earlier on a repayment plan than debtors who had received the standard letter.

### **Sample**

The required sample size was estimated prior to receiving the data using G\*Power (version 3.1.9.7; Faul et al., 2009) for one-tailed logistic regressions with a binary predictor. We used conventional alpha and beta values (.05 and .80 respectively) and assumed that the same number of debtors would receive the nudging and the standard repayment letter. The estimated effect size was based on pre-experimental data for the primary outcome (agreeing on a repayment plan) showing that a baseline of 19% of the debtors reached an agreement on an individual repayment plan, and an estimated increase of four percentage points (i.e., 23%). This increase was found as a lower threshold in similar experiments by the Behavioural Insights Team (BIT, 2012; Haynes et al., 2012; Service et al., 2014). This yielded an estimated effect size of  $OR = 1.27$  and a required sample size of  $N = 2,566$ .

Based on the number of repayment letters sent each month, we were confident to reach the required sample size given the researched periods. In fact, the data we received contained 3,568 cases in which a repayment letter had been sent. We excluded cases where a repayment letter had been sent to the same individual before ( $n = 188$ ), assuring that every individual appeared only once in our dataset. Further, we excluded cases in which the person had died before the outcome variables were measured ( $n = 37$ ), in which the debtor was reported younger than 18 years or older than 66 years ( $n = 9$ ), and in which the first payment was made on the same day the repayment letter had been sent ( $n = 4$ ), since in these cases payment could not be attributed to receiving that letter. This yielded a total sample size of 3,330 for the logistic regression. For the mediation analysis, we excluded cases where a payment took place before the letter confirming a repayment plan was sent ( $n = 81$ ) because agreeing was no meaningful mediator if it took place *after* repayment. It is plausible however that these cases resulted from debtors who transferred their first instalment immediately after agreeing on a repayment plan before the confirmation letter had been processed. Note that excluding these cases leads to a more conservative estimate of the indirect effect.

In Table 2.1, descriptive statistics and randomisation checks are depicted. Differences between the total sample and the subsample used to test the indirect effect were negligible. Compared to individuals receiving the standard letter, individuals receiving the nudging letter were slightly older according to Welch's test,  $t(3,294.3) = -3.15$ ,  $p = .002$ , and received fewer reminders according to chi-squared testing,  $\chi^2(1, n = 3,330) = 26.74$ ,  $p = .000$ .

**Table 2.1**

Descriptive statistics for the total sample by the letter received

	Standard letter ( <i>n</i> = 1718)	Nudging letter ( <i>n</i> = 1612)	<i>p</i>
Age	36.8 (10.8) <sup>a</sup>	38.0 (11.2) <sup>a</sup>	.002 <sup>b</sup>
Female	41% ( <i>n</i> = 696)	43% ( <i>n</i> = 696)	.128 <sup>c</sup>
Reminders sent	23% ( <i>n</i> = 401)	16% ( <i>n</i> = 260)	.000 <sup>c</sup>

Note: <sup>a</sup>*M* and (*SD*), <sup>b</sup> according to Welch's test, <sup>c</sup> according to chi-squared test.

For the regression discontinuity design, we used a sample of debtors who had received the repayment letter not longer than one month before (*n* = 405) or after the letter type had been changed (*n* = 278).

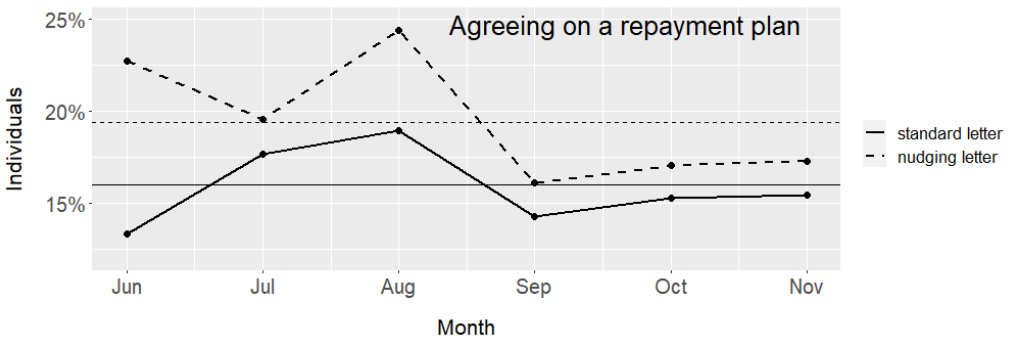
## Results

### *Agreement on a Repayment Plan*

Overall, 18% (*n* = 588) of all debtors reached an agreement on a repayment plan within 45 days after the repayment letter was sent. For individuals who had received the nudging letter, this percentage was higher (19%, *n* = 313) than for individuals who had received the standard letter (16%, *n* = 275). Compared to the null model, the model that included the letter type as a predictor provided better fit according to likelihood-ratio testing,  $\chi^2(1) = 6.65$ ,  $p = .001$ . However, predictive power of the model was low with Nagelkerke  $R^2 = 0.00$ .

**Figure 2.3**

Proportions of debtors agreeing on a repayment plan per month and letter type



Note: The horizontal lines indicate the means per letter type

In line with our theorising, we found a significant effect with debtors who received the nudging letter being 1.26 times more likely to reach an agreement on a repayment plan than debtors who received the standard letter,  $F(1, 3328) = 6.63$ ,  $p = .010$ . As a sensitivity analysis, we added gender

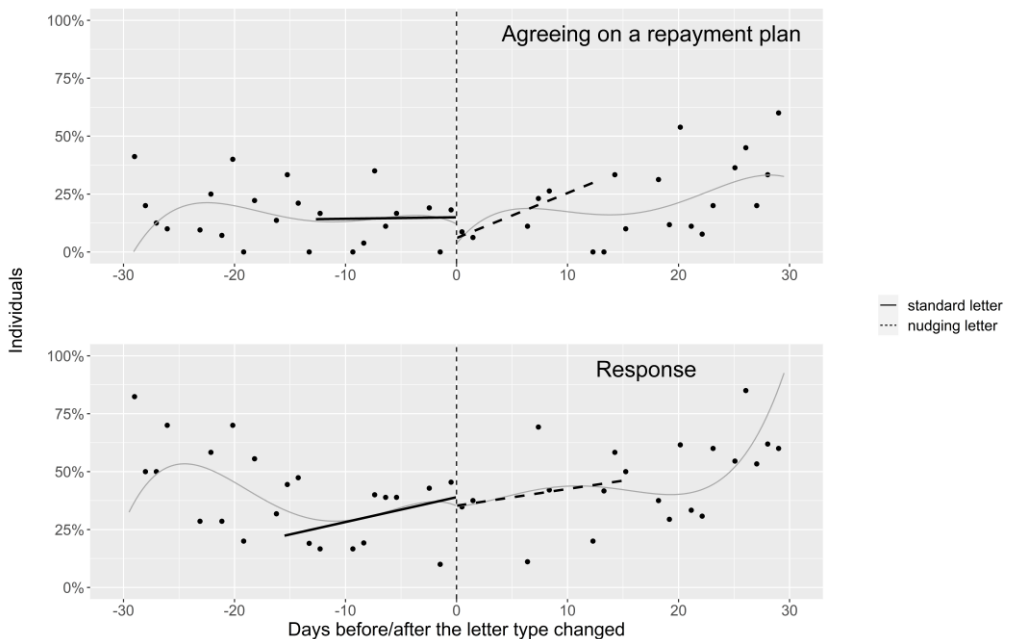
and age to the regression model. There was no evidence for multicollinearity. Without affecting the size and significance of the coefficient for the letter type, we found gender to be a significant predictor with women having agreed on a repayment plan more often (Table 2.2). In another sensitivity analysis, we inspected the proportion of debtors agreeing on a repayment plan each month (Figure 2.3). This proportion was consistently higher for the nudging letter when compared to the standard letter, implying a systematic increase due to the nudging letter rather than random variation between the two quasi-experimental periods.

The predicted probabilities to have reached an agreement on a repayment plan for a “normal” debtor (i.e., male, 37 years) was .14 with the standard letter and .18 with the nudging letter.

According to the regression discontinuity analyses, there was no significant local increase in debtors agreeing on a repayment plan when the letter changed,  $z = -0.93$ ,  $p = .354$ . However, the smoothed line in Figure 2.4 suggests an increase in early January and it is plausible that letters sent briefly before or during the holiday season led to irregular responses. Results did not change when covariates were included in the analysis.

**Figure 2.4**

Local change in outcomes at the day the letter type changed



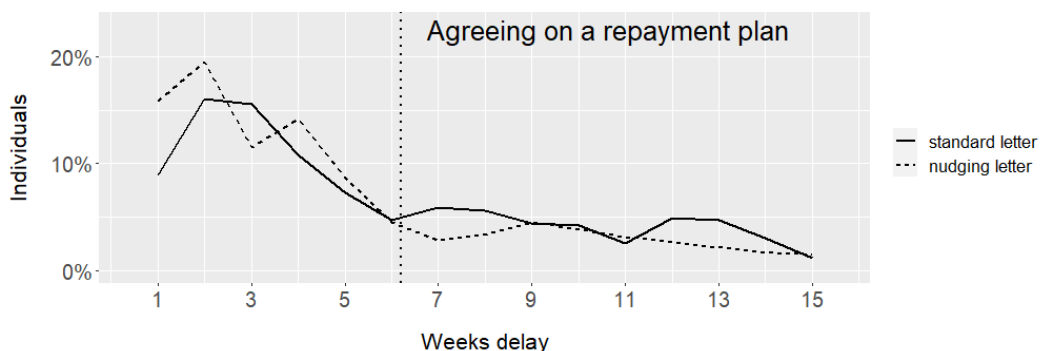
Note: The dashed vertical line indicates the day when the letter type was changed. The grey curved lines indicate smoothed proportions of debtors. Lines with slope indicate local regression lines.

Concerning the number of days that had passed between the repayment letter and the confirmations of a repayment plan, Figure 2.5 reveals that most debtors who agreed on a repayment

plan did so two or three weeks after the repayment letter had been sent. Different from the nudging letter, there is a peak for the standard letter after the deadline for a response to the repayment letter, possibly after having received an overdue notice. In general, debtors who agreed on a repayment plan did so earlier with the nudging letter ( $Mdn = 23$  days) than with the standard letter ( $Mdn = 28$  days), according to Wilcoxon signed-rank test,  $Z = -3.60$ ,  $p = .000$ .

**Figure 2.5**

Delay between the repayment letter and the confirmation of a repayment plan



Note: The dashed vertical line indicates the deadline for agreeing on a repayment plan that was stated in the repayment letter.

### Response

In total, 43% ( $n = 1,443$ ) of all debtors responded to the repayment letter before the six weeks' deadline either by agreeing on a repayment plan or by repayment. For individuals who had received the nudging letter, this percentage was higher (45%,  $n = 731$ ) than for individuals who had received the standard letter (41%,  $n = 712$ ). Model evaluation indicated better fit for the full model according to likelihood-ratio testing,  $\chi^2(1) = 5.16$ ,  $p = .023$ . However, predictive power of the model was again low with Nagelkerke  $R^2 = 0.00$ .

In line with our theorising, we found a significant effect with debtors who received the nudging letter being 1.17 times more likely to have responded to the repayment letter than debtors who received the standard letter,  $F(1, 3328) = 5.16$ ,  $p = .023$ . The sensitivity analysis did not affect the effect of the repayment letter but revealed that female and older debtors responded more often (Table 2.2). The predicted probabilities to have responded for a "normal" debtor were .38 with the standard letter and .41 with the nudging letter.

According to the regression discontinuity analyses, there was no significant local increase in debtors responding when the letter changed,  $z = -0.53$ ,  $p = .597$ , but again the smoothed line in Figure 2.4 suggests an increase in early January plausibly related to the holiday season. Results did not change when covariates were included in the analysis.



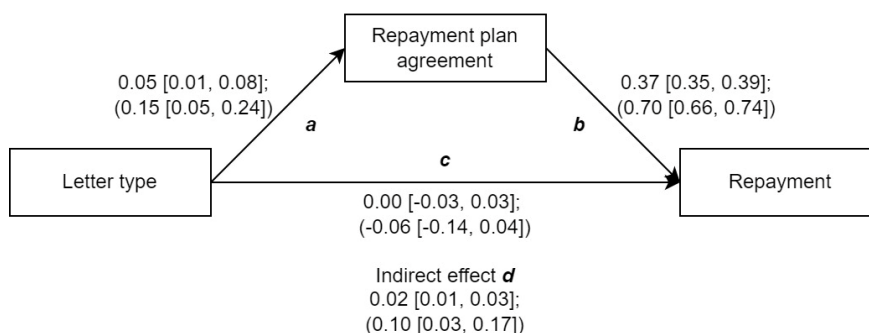
**Table 2.2**

Logistic regression results

Predictors	Agreement on a repayment plan		Response	
	OR	95% CI	OR	95% CI
Letter type (nudging letter)	1.26	[1.05, 1.50]	1.15	[0.00, 1.32]
Gender (female)	1.32	[1.11, 1.58]	1.44	[1.25, 1.66]
Age <sup>a</sup>	0.99	[0.91, 1.09]	1.11	[1.04, 1.19]
Constant	0.17	[0.14, 0.20]	0.61	[0.55, 0.68]

Note: OR = odds ratio, CI = confidence interval. <sup>a</sup>Standardised predictor.**Figure 2.6**

Standardised (unstandardised) coefficients and 95% percentile intervals for investigating the indirect effect

**Indirect Effect**

We found a significant indirect effect *d* (Figure 2.6) of the letter type on repayment via reaching an agreement on a repayment plan, with its 95% percentile interval for the standardized coefficient excluding zero [0.01, 0.03]. Individuals who received the nudging letter were thus more likely to have reached an agreement on a repayment plan and consequently have started repayment of the debt. The respective individual coefficients *a* and *b* were both significant with more individuals with the nudging letter agreeing on a repayment plan [0.01, 0.08] and more individuals with a repayment plan starting repayment [0.35, 0.39]. We found no evidence of a significant direct effect *c* of the letter type on repayment [-0.03, 0.03], indicating an indirect-only mediation (Zhao et al., 2010). This means that the letter type affected repayment only via agreements on a repayment plan. The findings remained the same when the mediation analysis was conducted with the covariates. Note that for some debtors, wage garnishment might have led to repayment by force rather than behaviour. However, this is implausible to have affected the effects *a*, *b*, and *d* because garnishment only took place after the response deadline and was delayed if debtors had agreed on a repayment plan. Nearly all debtors (92%) who agreed on a repayment plan had started repayment. All standardized and unstandardized effects are displayed in Figure 2.6.

## Discussion

In comparing the standard repayment letter of the municipality against the nudging letter, we found, in line with our hypothesizing, more debtors agreeing on a repayment plan after having received the nudging letter. In addition, overall response rates increased with the nudging letter meaning that the increase in debtors who agreed on a repayment plan was not driven by debtors avoiding immediate payment and agreeing on a repayment plan instead. Taking a long-term perspective, we found that agreeing on a repayment plan served as a mediator for repayment. However, the effects of the nudging letter were of small magnitude, increasing agreeing on a repayment plan by not more than three percentage points. Comparing these effects to other studies is complicated because of the unique setting and nudging techniques of this study. In fact, recent reviews have highlighted the context dependency of nudges and the heterogeneity of observed effects (DellaVigna & Linos, 2020; Hummel & Maedche, 2019). Yet, the effect sizes found here are twice those obtained by other BITs, but lower than those found in academic studies (DellaVigna & Linos, 2020). When compared to nudges aiming to increase tax compliance, the nudging letter achieved slightly higher effect sizes (Antinyan & Asatryan, 2020).

A lesson learned is that to achieve repayment in the context of financial scarcity, creditors should not focus on repayment but on contacting behaviour and agreeing on a repayment plan. In other contexts, contact might, however, not be necessary nor desirable since debtors with fewer liquidity constraints are more likely to be able to repay their full debts immediately without contact and agreeing on a repayment plan. In addition, we show that compared to earlier studies that adapted reminder letters for debt repayment (Janssen et al., 2017; Jensen et al., 2018), debt collection can be improved at low-cost earlier on by adapting debt repayment letters. Some studies suggest that larger and more robust effects sizes may have been obtained by including deterrence information in the nudging letter (e.g., specifying penalties for late payments; Antinyan & Asatryan, 2020; De Neve et al., 2021; Hallsworth, 2014). However, such information (and real costs incurred by debtors), would represent an additional stressor and burden for individuals experiencing financial scarcity potentially creating or reinforcing chronic indebtedness (Heidhues & Kőszegi, 2010; Lea, 2020).

The effects cannot be attributed to specific adaptations in the nudging letter because all adaptations were integrated in a single letter at once. For instance, we do not know if the longer text of the nudging letter increased cognitive demands thereby offsetting overall effectiveness of the letter. For future research it is hence of interest to test different adaptations and combinations thereof separately to specify their effects (Ludwig et al., 2011) or use non-experimental methods like surveys or interviews to help identify effective adaptations (Marchionni & Reijula, 2019). Combining multiple adaptations in single letters is common in applied settings, however (Janssen et al., 2017; Jensen et al., 2018).

Nevertheless, we can speculate about the effectiveness of different adaptations and underlying mechanisms. Many adaptations in the nudging letter can be categorised as simplifications (Münscher et al., 2016) reducing the cognitive effort necessary to process the information in the letter, namely

presenting choice options as bulleted lists, adding pictograms, and including also the concrete response deadline. Simplification has been identified as relatively powerful when compared to other nudging techniques (DellaVigna & Linos, 2020; Hummel & Maedche, 2019). However, given the short and already simple paragraph on contacting in the standard letter, it is plausible that not simplification, but the other adaptations were responsible for most of the observed effects, namely stressing togetherness and collaboration between debtors and creditors, improved opportunities for contacting, and adding action-relevant information. These adaptations reduced the administrative burden loaded on debtors receiving the standard letter (Moynihan et al., 2015). For instance, the social consequences of contacting were plausibly less threatening with the nudging letter because it freed debtors calling the municipality from the need to first explain their potentially stigmatising situation in sufficient detail to ensure being put through correctly by the general call centre staff. Instead, with the nudging letter, debtors could call the responsible staff directly. Reducing administrative burden was theorised to be especially effective for individuals experiencing (financial) scarcity (Sunstein, forthcoming), further supporting this reasoning.

Debtors who received the nudging letter seem to have agreed on a repayment plan earlier than debtors who received the standard letter. This is in line with similar nudge interventions used to increase tax compliance (De Neve et al., 2021). Shorter delays between the repayment letter and debtors' responses are practically relevant because they might lead to fewer reminders and overdue notices being sent, meaning that creditors can save costs and that debtors receive fewer potentially stressful communication. However, this finding still needs confirmatory testing to be able to draw valid conclusions.

Future research can innovate repayment letters further: Most adaptations in the nudging letter aimed to increase motivation or to make the target behaviour easier. Individuals however often fail to follow-up on their intentions even if they have the opportunity and if it is easy (Sheeran & Webb, 2016). Therefore, interventions aiming to improve the translation of intentions into behaviour, for instance implementation intentions (Gollwitzer, 1999) or enhanced active choice formats (Keller et al., 2011), may be added to future repayment letters. In addition, changing the envelope of letters may increase opening rates, further improving effectiveness. This may be done by printing importance appeals (e.g., "important information") on envelopes (Amos & Paswan, 2009). Further improving the letter is necessary because our results showed that the majority of debtors did not respond to it in time. However, a recent study showed that such avoidance behaviour is common when experiencing financial scarcity (Hilbert et al., 2022).

Like all research, this study is subject to limitations. As a practical limitation, instead of fixing the problem of unjustified welfare payments, it may have been better to develop and test measures avoiding those payments in the first place, for instance by stimulating welfare receivers to inform the municipality about additional incomes on time (e.g., Zhang et al., 2020). As a methodological limitation, the quasi-experimental design of this study means that our findings may be confounded by unobserved contextual factors (e.g., economic changes that affected incomes and how much individuals could repay) rather than the letter type. This seems to be the case for the regression

discontinuity analyses which plausibly was affected by the holiday season, highlighting the importance of choosing an adequate day for changing the letter with that type of analysis. Similarly, the differences might be attributed to random variance over time or regression to the mean. For future research it is of interest to fully randomise the letter to allow stronger conclusions.

As a final limitation, the data used for the analysis was not able to fully represent reality of the debt collection procedure because several relevant aspects were not registered as a standard. Plausibly the most relevant aspect was that municipality employees in rare cases proactively contacted debtors before the response deadline to suggest agreeing on a repayment plan. If more debtors receiving the nudging letter had been reached this way than debtors receiving the standard letter, the increase in debtors agreeing on a repayment plan might be the result of these contact attempts rather than the nudging letter. However, the public servants involved in debt collection had no reason to contact debtors more with the nudging letter and they themselves found it unlikely that this had been the case. Another important aspect of reality that we could not account for in the analysis is that not from all debtors a response was needed to repay the debt. This is because welfare receivers built up holiday allowance (i.e., money paid out once a year supposedly for going on holidays) and sometimes this allowance was used to fully repay the debt. This was booked internally by municipality employees, meaning that no action was required from the debtors. Based on pre-experimental data, this applies to approximately 7% of all debtors receiving the repayment letter.

As evidenced by these limitations, conducting field experiments is often difficult (Hansen & Tummers, 2020). Yet they are of high practical and scientific relevance because of their ecological validity. By accounting for the main positive outcomes of the debt collection procedure we believe that our results are useful for practitioners and scientists alike.

## Chapter 3

# Nudging is Ineffective When Attitudes Are Unsupportive

An Example From a Natural Field Experiment

This chapter has been published as:

Dewies, M., Schop-Etman, A., Rohde, K. I. M., & Denktas, S. (2021). Nudging is ineffective when attitudes are unsupportive: An example from a natural field experiment. *Basic and Applied Social Psychology*, 43(4), 213-225. <https://doi.org/10.1080/01973533.2021.1917412>

## Abstract

For security reasons, employees of a Dutch local government department needed to wear an identifying lanyard with their employee badge, but compliance with this policy was low. Two nudges to increase compliance were evaluated in a pre-registered natural field experiment using a pre-post design, and a qualitative survey. Bayesian inference provides insufficient support for the effectiveness of the nudges. While more respondents judged the nudges and the lanyard policy positively than negatively, there was substantial negative judgment and incomprehension for both with some employees finding the nudges paternalistic. We hypothesize that the nudges were ineffective because they failed to change attitudes about the policy, and because they led to reactance among some employees. Implications for nudging within organizations are discussed.

Government employees need to comply with policies for information security to protect sensitive information (e.g., addresses, health records, police operations) and mitigate risks of many sorts (e.g., privacy breaches, loss of trust, information loss). However, employees' compliance cannot be taken for granted. Indeed, employee compliance and a lack thereof is a major concern for information security (e.g., Hwang et al., 2017; Warkentin & Willison, 2009). Following Weaver (2014), compliance is dependent on three broad and related categories: perceived external incentives (incentives and sanctions, monitoring, enforcement), willingness to comply (information and cognition problems, peer effects, attitude and beliefs problems), and capacity to comply (resource problems, autonomy issues).

For this research, compliance with an information security policy at a Dutch local government department regularly dealing with sensitive information was investigated and we aimed to increase compliance by testing interventions in the field. This policy aimed to help identify unauthorized individuals at the department's office space. It required employees of that department to wear an identifiable lanyard with their employee badge around their neck when being present at the office space so that unauthorized individuals could be identified based on a missing lanyard. In aiming to increase compliance with that policy, the departmental management approached the Behavioral Insights Group Rotterdam (BIG'R; [www.bigrotterdam.nl](http://www.bigrotterdam.nl)). BIG'R, like other behavioral insights teams around the world (Afif et al., 2018), enables institutionalized collaborations between behavioral scientists and public servants to pioneer the application of behavioral insights for public policy (John, 2014).

The main barrier to compliance were willingness problems. Conversations with employees ( $N=8$ ) revealed that forgetting to wear the lanyard was a relevant cognition problem, that wearing it separated the employees from peers working for other departments who did not need to wear it, and that some employees' attitudes were unsupportive of the policy. The attitudes were unsupportive because employees did not believe in the effectiveness or necessity of the policy, for instance because they believed that they could recognize all employees even without the lanyard, or that the requirement to scan one's badge to get access to the department's office area was sufficient protection. As a result of this requirement, it was practical and often necessary for employees to carry the lanyard with them. Yet, they often had it in a pocket or bag, or held it in their hands instead of wearing it around their neck.

In developing interventions, BIG'R relied on nudging (Thaler & Sunstein, 2008) techniques. A nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, 2008). Nudges often exploit automatic cognitive processes (e.g., biases, heuristics) to stimulate behavior rather than engaging individuals in rational thought. The promises of nudges in the related field of safety compliance have been stressed elsewhere (Lindhout & Reniers, 2017) and reviews generally provide support for nudging techniques (Benartzi et al., 2017; DellaVigna & Linos, 2020; Hummel & Maedche, 2019). There is, however, little field research that examines nudging within organizations (Chapman et al., 2021). In addition, nudging research is in many ways still in its infancy and more evidence from the field is needed to determine when, how, for whom, and to what

extent nudging techniques work (e.g., Hummel & Maedche, 2019; Jachimowicz et al., 2019; Kusters & Van der Heijden, 2015; Lin et al., 2017; Marchiori et al., 2017).

We decided to trial multiple nudges to be able to address multiple barriers to compliance (i.e., forgetting and unsupportive attitudes; Weaver, 2015), and because the literature advises to take an integrated approach to compliance (via organizational culture) that could not be achieved with a single nudge alone (DeJoy, 2005; Neal et al., 2000; Sommestad et al., 2014). In an act of participatory research (Reason, 1994), ideas for nudges were brainstormed together with BIG'R employees and employees from the department ( $N=6$ ). Subsequently, the authors and BIG'R employees involved in this research discussed and selected the nudges to be tested. In this, we avoided nudges that would upgrade the lanyard (e.g., framing it as a perk) because this was assumed to further strengthen the perceived separation between the employees of the department and their peers from other departments who were not required or allowed to wear the lanyard.

The first nudge, here called head-start nudge after Thaler and Sunstein (2008), was a point-of-decision prompt that served to counteract forgetting to wear the lanyard. Point-of-decision prompts aim to disrupt habitual choices at the moment of making the choice and were hence believed to be well-suited to counteract forgetting. They were shown to affect, for instance, stair use (Nocon et al., 2010; Soler et al., 2010), healthy food choices (Cadario & Chandon, 2020), and hand hygiene (Caris et al., 2018; Weijers & de Koning, 2021). Specifically, stickers with an image of a person wearing the lanyard and text saying "You're holding it already, now just wear it" were placed at all printers and access points to the department's office space because there employees needed to scan their badges. The prompt exploited the insight that framing a task as begun but incomplete leads to an increased likelihood of task completion (Barasz et al., 2017; Nunes & Dreze, 2006; Ovsiankina, 1928). We assumed that with that reminder, employees would perceive scanning the badge and wearing it as part of a larger integrated task.

The second nudge, here called norm-awareness nudge, aimed to positively influence social norms because they were found to be an important predictor of compliance with information security policies (Bulgurcu et al., 2010; Guo et al., 2011) and because they are a powerful behavior change technique (Armitage & Conner, 2001; Cialdini & Goldstein, 2004; Miller & Prentice, 2016; Rhodes et al., 2020). Specifically, mirrors with a life-size print of the lanyard on it were placed at the department's office space. Observing oneself in a mirror is a typical manipulation to increase awareness of oneself and the nudge exploited the insight that awareness of oneself also raises awareness for social norms and has a positive effect on being faithful to those norms (Diener & Wallbom, 1976; Gibbons & Wright, 1983; Hofmann & Heinrichs, 2002; Wicklund, 1979). Placing mirrors together with an image of the lanyard we thus assumed would increase conformity with the injunctive norm to wear the lanyard. Moreover, we assumed that introducing the norm-awareness nudge after the head-start nudge would also increase awareness for a changing descriptive norm (i.e., an increasing number of employees wearing the lanyard as a result of the head-start nudge). Such increasing descriptive norms were recently found to be effective in promoting sustainable behavior (Loschelder et al., 2019; Sparkman & Walton, 2017). Moreover, a similar combination of social norm



information and mirrors was shown to have a positive effect on healthy food choices (Niculescu et al., 2016). By capitalizing on social norms, this nudge was hypothesized to counteract unsupportive attitudes.

Initially, the two nudges were planned to be complemented by a third nudge exploiting messenger effects (Pornpitakpan, 2004). Specifically, it was planned that new employees of the department would receive the lanyard from the department head during a departmental meeting instead of picking it up from an administrative employee. We assumed that this would signal management commitment which has been identified as a major determinant of safety performance (e.g., Christian et al., 2009; Zohar, 1980) and compliance with information security policies (e.g., Chan et al., 2005; Hu et al., 2012). Importantly, the effect of management commitment on compliance seems to be mediated by employee attitudes (Hu et al., 2012) which led us to hypothesize that this nudge would counteract unsupportive attitudes. However, the department head explained to us during the experiment that this nudge could not be carried out after he had received negative reactions from employees concerning the head-start nudge. The department head wanted to avoid more negative reactions and escalation. Following Weaver (2014), we theorized that the negative reactions were related to autonomy issues and decided to investigate why the nudges led to some negative reactions in Study 2 using a survey among the department's employees.

The guiding question for this research therefore was to what extent the nudges were able to affect compliance with the lanyard policy and how they were perceived. Approval for this research was obtained from the local ethics committee of the Department of Psychology, Education, and Child Studies at Erasmus University Rotterdam (approval number 19-040) and we obtained informed consent for Study 1 testing the effectiveness of the nudges from the head of the department because we collected data on department level, rather than individual level. For Study 2, we obtained informed consent from the respondents. Study 1 was registered online prior to inspecting the data (<https://osf.io/sdtf5>).

## Study 1

Study 1 served to evaluate the effectiveness of the nudges. For this, we conducted a natural field experiment at the department's office space and tested the following hypotheses. Note that these hypotheses have been rephrased in comparison to the preregistration document (e.g., deleting references to the messenger nudge) with their meaning unchanged.

H1: More employees will wear the lanyard correctly after implementation of the head-start nudge.

H2: More employees will wear the lanyard correctly after adding the norm-awareness nudge to the head-start nudge.

H3: The removal of all nudges, will not have an effect of the number of employees wearing their lanyard correctly at follow-up.

The department's office space was a secured area which only employees from that department could enter after scanning their badge. For the experiment, the department was considered the study population and the only participant.

**Methods**

*Context*

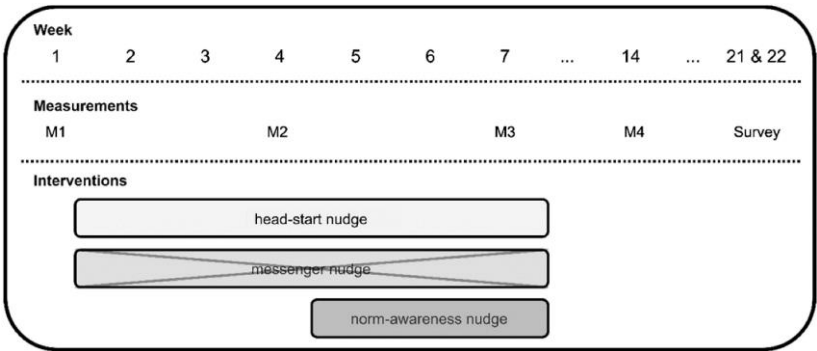
At the start of the study, 265 employees belonged to the department and had access to the secured area. The employees were on average 44 years old, the majority ( $n = 171$ , 65%) was female, and a small subgroup of employees ( $n = 13$ , 9%) had a supervisory function. The content of the employees' work regularly involved handling sensitive and personal information. Not all employees could work at the secured area at a time because only 137 individual workplaces were located there, mostly in one large open office space. No fixed workplaces were assigned within the secured area and if employees could not find a free workplace there, they could search for workplaces in a less secured office area next to the secured area. In this less secured area, employees were not required to wear the lanyard.

*Procedure*

The procedure is summarized in Figure 3.1. Measurement 1 (M1) in week 1 (May 6 to 10, 2019) served as a pre-measurement of the proportion of employees wearing their lanyard. In week 2, stickers with the head-start nudge were placed at the two printers and the nine badge scanners at the entries to the secured working area. Starting in week 4, the first post-measurement (M2) was carried out to evaluate the head-start nudge. Because of a bank holiday in week 4, M2 also included the first day of week 5. After M2, four mirrors with the lanyard on it were placed (week 5). In week 7, the second post-measurement (M3) was carried out to measure the incremental effect of the norm-awareness nudge. After M3, both nudges were removed and a follow-up measurement (M4) was conducted in week 14. Note again that the messenger nudge displayed in Figure 3.1 was not executed for reasons explained above.

**Figure 3.1**

Schematic display of the study procedure



### *Measures*

Compliance with the lanyard policy was measured by two human counters who walked the secured working area on a specified route during standard working hours using handheld tally counters. Before the start of the study, all counters received instructions and training. The first counter, a research assistant conducting all measurements, counted all individuals present and those wearing the lanyard around their necks. Individuals wearing a uniform of support staff (e.g., cleaning personnel) were excluded from being counted. As supervisors could not be recognized by the research assistant, differing municipality employees accompanied the research assistant to count the supervisors present and those wearing the lanyard around their necks. Supervisors thus form a subgroup of the counts of all employees.

Each measurement encompassed ten counting instances during five consecutive working days. For each measurement the counters walked the secured area twice a day with a delay of minimum 2.5 hours and maximum 6 hours. It was not possible to rule out that some employees would be counted multiple times during a single counting instance (i.e., when an employee moved from an area already walked by the counters to an area not yet walked). However, the chance that more than a few employees were counted multiple times was low as the employees mostly conducted sedentary deskwork (hence little movement between workplaces) and counting instances lasted maximum five minutes. In addition, counters had a good oversight over the secured area and could notice when employees moved. Whenever someone had questions about the purpose of the counting, the counters told a cover story about the occupancy rate of the office area. For two counting instances from each measurement, the first author accompanied the two counters acting as a parallel counter to the research assistant. This allowed us to calculate the reliability of the counting method as Krippendorff's alpha (Krippendorff, 2004) with an excellent value of .98. Krippendorff's alpha is a common measure in content analysis for the extent of agreement between coders when coding unstructured data or observations.

### *Analytic Strategy*

The effectiveness of the nudges was evaluated for all employees including supervisors. We started the analysis with the detection of possible outliers in the compliance percentages. Outliers were empirically predefined as counting instances where the z-score of the percentage was larger than 2.58 (i.e., outside a 95% confidence interval for normally distributed data).

To investigate the effects of the nudges, we relied on Bayesian modeling which is often more flexible compared to standard (frequentist) procedures (van de Schoot & Depaoli, 2014), thereby allowing us to test all hypotheses within one statistical framework that made use of all statistical information. The aim of this modeling was to generate posterior distributions through MCMC chains that integrated priors with the observed data. These posterior distributions were then used to infer point estimates for modeled parameters and intervals of interest. To do this, we relied on the model structure described in Kruschke (2015, pp. 251–260) and adapted the prior specification in the accompanying R code. We specified vaguely informative priors which meant that the posteriors (and

thus our results) were almost completely informed by the observed data and only marginally by prior beliefs. The vaguely informative priors meant that compliance values very close to 0% or 100% were believed to be unlikely as full (in)compliance is rare. We provide a description of the model, the model code, and details concerning initialization of the MCMC chains in the online supplementary material (<https://doi.org/10.1080/01973533.2021.1917412>).

The primary interest of our analyses were the different proportions of employees complying with the lanyard policy during the four measurements. In the model structure, these proportions were represented by the  $\omega_c$  parameters with  $c$  indicating the measurement (e.g.,  $\omega_{M1}$  represents the proportion at measurement 1). They were defined as the modes (i.e., the most likely points) of the posterior distributions describing the compliance rate for measurement  $c$ . Subtracting the posterior distributions of compliance rates during different measurements from each other yielded new posterior distributions that described the change in the proportion of employees complying with the policy between measurements. Investigating these new distributions allowed us to inspect the effectiveness of the nudges. Specifically, we investigated the modes of these differences (denoted  $\omega_{c-c'}$ ) as point estimates for the change in compliance rates between measurements, and the 95% highest density intervals (HDI) of these distributions which included 95% of the most credible values around that mode. If the 95% HDIs did not include the null value, a change in compliance between measurements was inferred.

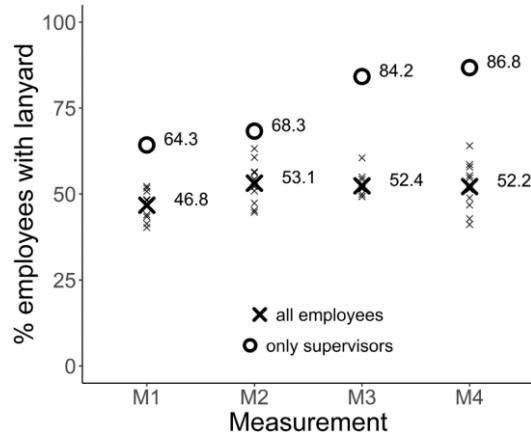
The analysis was conducted using R version 4.0.2 and JAGS. After initial observation of the MCMC chains (van Ravenzwaaij et al., 2018), we specified 1,000 adaptation steps, 200 burn-in steps, and 500,000 saved steps. Convergence of the chains was evaluated by ensuring that the value of the Gelman-Rubin criterion (Gelman & Rubin, 1992) was smaller than 1.01 for all parameters (Vehtari et al., 2021) and by visually inspecting the trace plots for all  $\omega_c$  parameters. In addition, we ensured that the effective sample sizes (ESS) of the posterior distributions describing the compliance rates per measurement (from which the  $\omega_c$  parameters were inferred) and the differences between these distributions (from which the  $\omega_{c-c'}$  parameters were inferred) exceeded 10,000 which has been recommended as a minimum to obtain stable limits of 95% HDIs (Kruschke, 2015). This also applies to the tails of all these posterior distributions which is important to achieve precision for the limits of 95% HDIs (Vehtari et al., 2021).

## Results

In total, individual employees were counted 2,246 times during the 40 counting instances. This meant that on average 56.15 ( $SD = 18.17$ ) employees were counted per counting instance. Inspecting the percentages of employees wearing their lanyard, we found no evidence of outliers. Figure 3.2 displays the percentages of employees complying with the lanyard policy per measurement. At M1, we measured 47% of employees wearing their lanyard averaged across counting instances. From M1 to M2 there was an increase of six percentage points. Thereafter only subtle changes were observed. For supervisors, there was a step-wise increase of 23 percentage points from M1 to M4.

**Figure 3.2**

Compliance per measurement



Note: Percentage of employees wearing their lanyard per measurement (bold) and per counting instance (non-bold). Supervisors form a subgroup of employees. Due to the small number of supervisors per counting instance, only the average over all counting instances is displayed for them.

**Table 3.1**Summary statistics for the  $\omega_c$  and  $\omega_{c-c'}$  parameters

Parameter	Mode	ESS	95% HDI	
			low	high
$\omega_{M1}$	.470	189,382	.415	.525
$\omega_{M2}$	.537	238,188	.477	.596
$\omega_{M3}$	.525	271,881	.472	.577
$\omega_{M4}$	.530	309,958	.469	.588
$\omega_{M2} - \omega_{M1}$	.066	181,538	-.014	.147
$\omega_{M3} - \omega_{M2}$	-.012	279,745	-.090	.066
$\omega_{M4} - \omega_{M3}$	.002	310,351	-.074	.081

Note: ESS = effective sample size.

Turning to inference statistics and the results from Bayesian modeling, the diagnostic criteria indicated that all MCMC chains seemed to have converged. In Table 3.1, the summary statistics for the  $\omega_c$  and the  $\omega_{c-c'}$  parameters are displayed. Note that the modes did not differ from the means of the posterior distributions by more than 0.002. Inspecting the 95% HDI for  $\omega_{M2} - \omega_{M1}$ , we did not find sufficient support of an increase in compliance from M1 to M2 as the interval  $[-.014, .147]$  includes the null value. Hence, we do not infer an effect of the head-start nudge. The same held when inspecting the 95% HDI for  $\omega_{M3} - \omega_{M2}$  as it also includes the null value  $[-.090, .066]$ . Consequently, we do not infer an effect of the norm-awareness nudge when added to the head-start nudge. There was also no

effect of removing the nudges as the 95% HDI of  $\omega_{M4} - \omega_{M3}$  did include the null value  $[-.074, .081]$ . Note that the conclusions do not change when supervisors are excluded from the analysis.

During the experiment, the department head received some negative reactions from employees who were irritated by the head-start nudge because they found it “too much.” In addition, the head-start nudges themselves received some negative reactions, for instance when the image of a model was glued to the sticker to cover the image of the person wearing the lanyard and it seemed like employees had removed some of the nudges (which were then replaced).

## ***Discussion***

This study provides insufficient support for the effectiveness of the nudges. There is no support for hypotheses 1 and 2. There is support for hypothesis 3 as the removal of the nudges did not affect compliance. During the experiment, one of the nudges was not executed because of negative reactions from employees concerning the head-start nudge. We could only speculate about the underlying reasons for these reactions. It seemed plausible, however, that autonomy issues (Weaver, 2014) played a role as nudges have been criticized for their paternalistic notion (e.g., Mitchell, 2004), and because the negative reactions could be interpreted as an attempt to restore one's autonomy according to reactance theory (Brehm, 1966; Brehm & Brehm, 1981). Therefore, we decided to investigate the perception of the nudges in a second study.

## **Study 2**

Study 2 encompassed an exploratory, cross-sectional survey amongst the department's employees that served to investigate why the nudges led to some negative reactions and why the nudges were ineffective. In combining quantitative and qualitative methods, we endorsed a pragmatist approach to research (Johnson & Onwuegbuzie, 2004; Morgan, 2007).

## ***Methods***

### ***Study Population***

Employees who had worked at the secured working area at least once a week during the experimental period were considered to belong to the study population because they could be assumed to have experienced the nudges. As we do not know for how many employees this requirement was fulfilled, we also do not know the exact size of the study population. In total, 142 employees agreed to participate and answered the survey (54% of all employees of the department). See Study 1 for the demographics of all employees.

### ***Procedure***

Seven weeks after the follow-up measurement (M4; see Figure 3.1), all employees of the department were invited via email by the department head to fill in the online survey. They were given two weeks' time to fill in the survey, with a reminder sent after one week. Before employees could fill in the survey,

they were asked to confirm that they had worked at the secured area where the experiment had been conducted at least once a week during the experimental period.

### *Measures*

**Compliance:** As a self-reported compliance measure, respondents were asked how often they were wearing the lanyard at that moment on a scale from 1 (never) to 7 (always).

**Lanyard policy:** Respondents were asked what they thought about the lanyard policy using an open answer format.

**Nudges:** Respondents were asked separately for both nudges what they thought about them using an open answer format. Respondents were only asked this question for the nudge(s) that they remembered.

**Paternalism of the nudges:** Respondents were asked separately for both nudges to what extent they found them paternalistic on a scale from 1 (not at all paternalistic) to 7 (totally paternalistic). In addition, respondents were asked to explain their answer(s) using an open answer format, again, separately for both nudges. Respondents were asked these questions after they reported their general thoughts on the nudges and only for the nudge(s) that they remembered.

**Counting instances:** Respondents were asked using an open answer format what they thought about the counting instances.

**Purpose of the counting instances:** Using open answer formats, respondents were asked separately what they thought was the purpose of the counting instances, and if the purpose they assumed had changed some time after the first counting instance.

**Reasons for negative reactions:** Using an open answer format, respondents were asked what they thought were the reasons for some negative reactions to the nudges.

**Improving compliance:** Using an open answer format, respondents were asked to think of aspects that would increase their motivation to wear the lanyard.

All survey questions are provided in the online supplementary material (<https://doi.org/10.1080/01973533.2021.1917412>).

### *Analytic Strategy*

The answers to closed questions were analyzed using descriptive statistics and visual displays. The answers to open questions were analyzed using both qualitative coding and descriptive statistics: In a first step, the first author and a municipality researcher independently familiarized with the data and coded the answers from a subset of 15 respondents in Atlas.ti 8 using the constant comparison method (Boeije, 2002). In a second step, those codes and the emerging codebook were discussed and revised. The first author then used the resulting codebook to code the answers from all respondents in a third step and validated ambiguous quotes with the municipality researcher. Later additions to the codebook were made in agreement between the two coders.

The coded answers were used to generate a code-respondent table that showed for each respondent (as rows) if a specific code (as columns) had been applied to the respondent's answer. Note that this table was based on a code-document table generated in Atlas.ti 8 where answers from each respondent were stored in different documents. Consider a hypothetical example: The answer from respondent K elaborating on a nudge "I found it an eye-catcher" may be assigned the code "salient." The cell value belonging to respondent K and the "salient" code would then be 1. It would be 0 if the "salient" code had not been applied. Counting the number of respondents that had provided answers related to the "salient" code and comparing it to the total number of respondents would then allow us to examine the prevalence of the code. With the "salient" code being applied to the answers of 15 respondents (i.e., the sum of the values in the "salient" column), the prevalence of this code would be  $15/142 = 11\%$ . The prevalence of different codes was compared for all codes belonging to the same measure.

Note that prevalence is a feature of an individual code and that prevalence provides no information on the overall distribution of codes across respondents. As an example, if the codes "salient" and "dull" both were to have a prevalence of 50% this would not imply that half of the respondents provided answers related to the "salient" code and the other half answers related to the "dull" code. In fact, the same half could have provided answers to which both codes were applied. However, our code-respondent table allowed us to investigate the co-occurrence of codes (i.e., whether codes tended to be assigned to the answers from the same or different respondents).

## Results

In terms of self-reported *compliance*, the majority of respondents indicated to always or nearly always wear their lanyard (Figure 3.3). At the same time, some respondents (15%) reported to never wear their lanyard. Inspecting the prevalence of the codes concerning the *lanyard policy*, we found that more than one third found the lanyard policy "fine" (Table 3.2), which in some cases included positive judgements such as "I find it nice." More than one fourth of the respondents expressed incomprehension concerning the policy. This was mainly because they believed that they could recognize all employees without the lanyard, or that the requirement to scan one's badge to get access to the office area of the department was sufficient protection. In contrast to that, a group of similar size did express comprehension for the policy and its underlying motivation of being recognizable. Nearly one fifth of the respondents expressed a negative judgment of the policy, for instance they found wearing the lanyard "not pleasant." 15% of the respondents found that visibly wearing the lanyard in any way, rather than around one's neck, would suffice. Finally, roughly one out of ten respondents stated that the policy put employees of the department in a separate position compared to their peers from other departments who were not required to wear any identifiable objects.

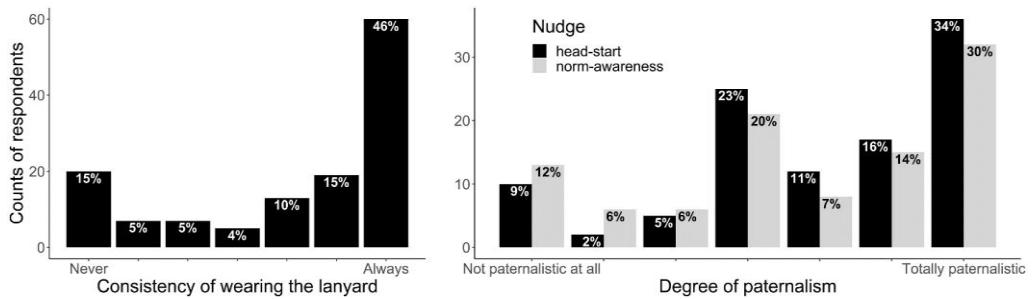
Turning to the head-start and the norm-awareness nudge, a large majority of respondents remembered to have seen them (84% and 82% respectively). Importantly, the answers from the open-ended questions did not differ much between the two *nudges* and respondents often copied their



answers or referred to answers given concerning the other nudge. We therefore did not differentiate between the two nudges in the analysis.

**Figure 3.3**

Answer distributions for self-reported compliance (left) and paternalism ratings of the nudges (right).



More than one third of the respondents found the nudges fine or positive. This code has a double label because “fine” (Dutch “prima”) can be interpreted as “satisfactory” as well as “good” and we could, in many cases, not differentiate what specific meaning was intended. Close to one third of the respondents expressed negative judgements of the nudges, sometimes quite strongly. Nearly one fifth showed little or no comprehension for the nudges, for instance stating that they did not need a reminder, or that being informed once concerning the policy suffices. The nudges were found some sort of paternalistic by 18% of the respondents. Note that these answers likely were provided before respondents read the question that asked them to rate the *paternalism of the nudges*. Inspecting the answer distributions for those ratings, most respondents found the nudges paternalistic to some extent at minimum (Figure 3.3). Asked to explain their paternalistic ratings, respondents provided answers related to the codes of incomprehension, negative judgment, and paternalistic aspects (Table 3.2). In addition, more than 10% found the nudges paternalistic because they perceived them as childish and reflecting an unjust treatment of employees.

We report the results for the remaining measures in less detail because fewer respondents answered the related questions. In addition, we judged the results to overlap in part with the results reported above or to confirm earlier research findings, thereby providing little additional insight. For instance, answers about *reasons for negative reactions* overlapped to a large extent with what many respondents thought of the lanyard policy and the nudges (e.g., many respondents finding the nudges paternalistic using this to explain the negative reactions). In addition, these answers often seemed to be speculative (e.g., many respondents indicated that they were speculating about their peers’ motives). Concerning the *counting instances*, the answers generally reflected a discomfort of being a research subject, thereby confirming earlier research findings discussed elsewhere (Jones & Whitehead, 2018). 12% of the respondents indicated that they assumed or knew that the *purpose of the counting instances* was to count employees wearing the lanyard. The most prevalent suggestion for *improving compliance* was made by 23% of the respondents and was to generate a better

understanding for the lanyard policy. Tables with the full results for these measures can be found in the online supplementary material (<https://doi.org/10.1080/01973533.2021.1917412>).

**Table 3.2**

Codes related to the lanyard policy and the nudges with a minimum prevalence of 10%

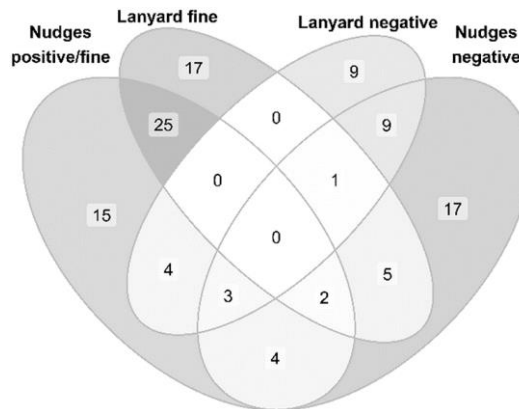
Prevalence	Code	Quotes (translated)
<i>Lanyard policy</i>		
35% (50)	Fine	"fine", "good", "I find it nice"
27% (39)	Incomprehension	"I don't find it necessary", "We already occupy a locked area where it is difficult to enter"
25% (36)	Comprehension	"I have comprehension for it", "good for recognizability"
18% (26)	Negative	"terrible", "not pleasant", "going too far"
15% (22)	Visibly suffices	"Wearing it well visibly seems okay to me as well"
11% (15)	Separate position	"It's strange that it only applies to us", "It's in the way of connecting with other branches"
<i>Nudges</i>		
37% (53)	Positive/fine	"fine", "playful", "clear", "fine, good stimulation"
29% (41)	Negative	"irritating", "a bit too much", "annoying"
18% (25)	Incomprehension	"unnecessary", "I think a message about it is already enough"
16% (23)	Paternalistic	"patronizing", "authoritarian", "pedantic"
10% (14)	Mirror positive	"I find the mirrors useful though"
<i>Explanation for the paternalism ratings</i>		
16% (23)	Incomprehension	"I don't need a sticker to be reminded to wear it"
14% (20)	Negative	"Too much of a 'hassle'", "The message was conveyed too often and to prominently", "language was not nice"
13% (19)	Childish	"childish", "elementary-school-like", "We are professionals. Treat me accordingly"
11% (16)	Paternalistic	"patronizing", "authoritarian", "pedantic"
11% (16)	Unjust treatment	"Seems like you don't dare to talk to each other anymore", "Seems like that employees of [...] can't think for themselves anymore"

Inspecting the *co-occurrence* of codes, we found that the evaluative judgment of the lanyard policy corresponded with the judgment of the nudges. Respondents who judged the lanyard policy as fine or positively also tended to judge the nudges positively as shown in the Venn diagram in Figure 3.4. Both codes together were assigned to 27 respondents, which translates to 19% of all respondents. If both codes had been assigned to respondents randomly at the same base rates, one would expect 19 (13%) respondents with both codes assigned together. Moreover, there was little

overlap between positive and negative judgements, implying that they form different clusters. In fact, only 19 respondents provided answers to which both positive codes (concerning either the lanyard policy or the nudges) and negative codes (concerning either the lanyard policy or the nudges) were assigned. This translates to 13% of all respondents. Again, if codes had been assigned to respondents randomly at the same base rates, one would expect 28 (20%) respondents to be assigned both positive and negative codes.

**Figure 3.4**

Co-occurrence of codes related to the nudges and the lanyard policy



### Discussion

The results show that after the experiment unsupportive attitudes concerning the lanyard policy prevailed among a substantial subgroup of employees. The tendency of employees to judge both the lanyard policy and the nudges negatively seems to imply that negative reactions did not stem from the nudges alone but plausibly from a combination of thinking negatively about both. This reasoning, however, only applies to a subgroup of employees as the most prevalent codes concerning the lanyard policy (and the nudges) were positive or at least neutral.

### General discussion

This study aimed to investigate the effectiveness of the head-start nudge and the norm-awareness nudge in increasing compliance with a security policy requiring employees to wear a lanyard with their employee badge attached to it. We found insufficient support for the effectiveness of the nudges (hypotheses 1 and 2). Unfortunately, it is common that behavior change interventions have no effect (Osman et al., 2020). It is therefore important to investigate factors that likely influence effectiveness. In doing so, a survey among the department's employees revealed that more respondents judged the nudges and the policy positively rather than negatively. Yet, the nudges were judged negatively by a substantial part of the employees and some found them unnecessary or paternalistic. Despite the

lower prevalence of negative judgements, the discussion will elaborate more on those because in our opinion they lead to interesting and relevant insights.

We hypothesize that employees holding attitudes that were unsupportive of the lanyard policy limited the effectiveness of the nudges, particularly the norm-awareness nudge. It has been argued that unsupportive attitudes can cause nudges to be ineffective because of the choice preserving nature of nudges (de Wijk et al., 2016; Sunstein, 2017). Individuals then follow their attitudes rather than the nudge. Interviews before the experiment and the survey both show that employees did not believe in the policy's effectiveness or necessity (code "incomprehension"). The nudges aimed to change attitudes in a subtle way, mostly by raising awareness for social norms. Following norm activation theory (Schwartz, 1977; Yazdanmehr & Wang, 2016), however, the injunctive social norm may not have led to attitude change and associated compliant behaviour because defense mechanisms allowed employees to neutralize the obligation to wear the lanyard. Specifically, it may be difficult to comply with the policy when there is a lack of personal consequences (e.g., sanctions, incentives) and the effectiveness or necessity of the lanyard policy were not comprehended. From that perspective, it follows that the nudges cannot be concluded to be ineffective per se but that they were ineffective *in the specific context* of this study. It remains an open question if the nudges had been more effective if they had been backed up by explanation about the lanyard policy as suggested by some employees, or if more rationality enabling nudges had been employed as sometimes suggested in the literature (John et al., 2009; Tor, 2020).

Similarly, we can only speculate about how the outcomes of the experiment might have differed if the messenger nudge had been executed. The department head stressing his commitment to the lanyard policy, especially if accompanied with explanation for why the policy was needed and how it was assumed to contribute to information security, might have given the lanyard policy and the nudges more credibility and authority, thereby potentially increasing effectiveness of the nudges.

We hypothesize that the negative reactions following the nudges were the result of autonomy issues experienced by some employees (Weaver, 2014). In fact, the employed nudges may have been less choice preserving than what is typically assumed for nudges (Thaler & Sunstein, 2003; 2008) as complying with the policy was the only acceptable response. From this perspective, the nudges were reminders of a mandatory policy and hence shoves rather than subtle features of a choice context. Consistent with this, some of the employees found the nudges paternalistic. The negative reactions as well as not complying can accordingly be interpreted as attempts to restore one's perceived autonomy whether to wear the lanyard or not (Brehm, 1966; Brehm & Brehm, 1981). This effect was plausibly reinforced by some employees associating the counting instances with the lanyard policy and feeling uncomfortable about being observed. Note however that nudges were found to be effective in other regulated contexts where nudges also might be considered shoves (e.g., Hallsworth et al., 2017; Rogers & Feller, 2018; Wu & Paluck, 2021).

It is plausible that some negative reactions could have been prevented if there had been an opportunity for employees to voice their concerns about the lanyard policy in a constructive manner instead (e.g., by stating a contact person on the nudges) as it is also indicated by the survey results:

Some employees felt being treated like children (code “childish”) and some employees felt a lack of open communication (code “unjust treatment”). This points to the importance of interactional justice (Colquitt et al., 2005) in organizational settings for nudges to work.

It is not consistent that Dutch public servants who in general tend to embrace the application of behavioural insights, reacted negatively when they themselves were the subject of the application. In the Netherlands, behavioural insights are popular and have been embraced by various government institutions such as the Dutch Scientific Council for Government Policy (Wetenschappelijke Raad voor het Regeringsbeleid, 2014). Somewhat inconsistent with that, negative perceptions of the nudges were of high prevalence among the employees (codes “negative,” “incomprehension,” “paternalistic”). This inconsistency can be intra-personal and inter-personal: The former implies that public servants endorse applying behavioural insights but dislike being the subject of that application and the latter suggests that there may be a divide between public servants who generally approve the application and those who reject it.

Like all research conducted, this study is not without limitations. First, the pre-post design limits the potential to draw any causal conclusions. In the given context, a more rigorous randomized control trial was not possible. Moreover, the context of this research with its specific mix of preexisting preferences concerning the target behavior limits generalizability claims. However, in adopting a case study perspective, this study yields valuable hypotheses to be tested in future research. For instance, that nudges perceived as childish will lead to feelings of reactance and noncompliance.

Another limitation is the delay of 14 weeks between the end of the intervention period and the start of the survey. Although approximately four out of five respondents indicated to remember the nudges, recalling immediate and original reactions to the nudges might have been difficult due to decaying or distorted memories. As a result, answers might have been affected by talks that employees had with other employees during the delay or knowledge concerning nudging they acquired in the meantime.

Another important limitation was that to conduct this research, some employees needed to be informed about it (e.g., the department head). In fact, 12% of the employees indicated to associate the counting instances with the lanyard policy which means that some additional employees had inferred the purpose of the counting instances from the behavior of the counters. We find it unlikely though that this led to socially desirable behavior (i.e., wearing the lanyard) because there were no incentives or sanctions conditioned on the employees’ behavior and because it was unpredictable for employees when counting instances would take place. In fact, we observed that the employees who needed to be informed did not always wear their lanyard either. Note that relatively more supervisors than employees needed to be informed about the experiment and that this might explain the larger increases in compliance among that group.

A strength of this study is its reliance on both quantitative and qualitative methods to evaluate the nudges beyond their effectiveness. The effectiveness measure and the survey represent measurements on different levels and from different perspectives: the former reflects an objective group characteristic and the latter subjective beliefs and judgements of individuals. One should keep

in mind, however, that individual judgements are conceptually limited when explaining group characteristics without investigating translational processes (e.g., interactions related to the lanyard between employees, differential treatment effects). Nevertheless, the survey enabled us to evaluate the context more thoroughly investigating limiting conditions for nudges to work.

We conclude with practical implications for field experiments involving nudges and suggestions for future research. Concerning the first, we advise researchers as well as practitioners to survey and reflect on the target group's attitudes and preferences before the start of an experiment testing nudges and we advise to inform the target group as much as possible about the nature of the planned experiment to anticipate potential negative reactions. In the case of unsupportive attitudes researchers may consider rationality enabling nudges or other behavior change techniques. For future research, we suggest more research into nudging within organizations which, as this research exemplifies, are a challenging environment for nudging interventions. It is also of interest how nudging in such contexts affects or hurts relational concepts such as trust and organizational citizenship behavior.

## Chapter 4

### Committing to Keep Clean

**Nudging Complements Standard Policy Measures to Reduce Illegal Urban Garbage Disposal in a Neighborhood With High Levels of Social Cohesion**

This chapter has been published as:

Merkelbach, I., Dewies, M., & Denктаş, S. (2021). Committing to keep clean: Nudging complements standard policy measures to reduce illegal urban garbage disposal in a neighborhood with high levels of social cohesion. *Frontiers in Psychology*, 12, 660410. <https://doi.org/10.3389/fpsyg.2021.660410>

## Abstract

Illegal garbage disposals are a persistent urban problem, resulting in high clean-up costs, nuisance and decreased satisfaction with the neighborhood among residents. We compared three adjacent city-areas in Rotterdam in the Netherlands where, for 2 weeks, either: (1) no action to decrease illegal garbage disposals was taken; (2) standard door-to-door canvassing was carried out; or (3) door-to-door canvassing was enriched with several nudges, most importantly a commitment nudge. The nudge treatment proved highly effective, reducing illegal disposals at post-test and follow-up (2 months later) with two-thirds, resulting in a very large effect size ( $d = 2.60$ ). At post-test, standard door-to-door canvassing did not differ from the control treatment, but at follow-up results were comparable to the nudging-treatment. This could, however, be due to spill-over effects. Using a commitment nudge thus proved highly effective in decreasing illegal garbage disposals, however, effects might be specific to neighborhoods with strong social cohesion.



Illegal garbage disposal is a serious problem in Rotterdam, the second largest city of the Netherlands with approximately 625,000 inhabitants. In a representative survey among the city population (Neighborhood Profile Rotterdam, 2016), 19% named littering as their number one nuisance, compared to 10% on national level. Concerning the consequences of disposing garbage in public, littering was shown to reduce the aesthetical quality of the environment (Roda et al., 2016). Brown and Raymond (2006) showed that residents identify aesthetical quality as very important, and that it was predictive for residents' place attachment. Place attachment in turn was found to be positively related to life satisfaction, further highlighting the importance of reducing illegal garbage disposal to improve the subjective experiences of Rotterdam residents (Casakin & Reizer, 2017). Moreover, reduced place attachment is related to reduced efforts into caring for one's residential environment (e.g., Mohapatra & Mohamed, 2013).

Littering is a potentially self-reinforcing problem. Various studies have shown that visible garbage in public spaces invites individuals to litter more in these spaces themselves (e.g., Keizer et al., 2011; Schultz et al., 2013). As an explanation it has been suggested that the presence of garbage indicates the prevailing social norm for how to dispose one's garbage, i.e., by leaving it in public spaces (Cialdini et al., 1990). Garbage placed next, instead of into, containers is a huge problem.

To prevent the emergence of a vicious littering-cycle as well as to increase residents' attachment with their neighborhood, in the past the municipality of Rotterdam put substantial effort into keeping public spaces clean. Illegal garbage was for instance frequently picked up by the municipality, up to multiple times a day in the neighborhoods with the most severe littering problems. This was, however, very costly and did not reduce illegal disposals, possibly because it stimulated free riding behavior on public services (Fischbacher & Gächter, 2010). Until now, no effective interventions have been implemented in Rotterdam. A recent experiment in which pick-ups were reduced in an attempt to increase residents' own responsibility did also not have the desired effect (Dur & Vollaard, 2015), but even resulted in increased littering. Because current practice (i.e., frequent pick-ups) is costly and reduced pick-ups resulted in more littering, we focused on measures beyond pick-up frequency.

Specifically, we aimed to improve the standard canvassing policy of the municipality adding to it a "behavioral spin" (Loer, 2019). Typically, policy measures assume individuals to react to them rationally (Howlett, 2018) meaning that they are designed to target rational thought. In fact, canvassing focused on information about rules for and possible consequences of illegal garbage disposal (e.g., fines, attracting vermin) hoping that information would stimulate behavior change. However, rational-based approaches regularly fall short limiting the effectiveness of policy measures (Weaver, 2015). Therefore, it has been suggested to use behavioral insights to improve effectiveness (Thaler & Sunstein, 2008). A challenge, however, has been how to complement existing measures with behavioral insights (Loer, 2019). Such complementary measures have recently been described as the "most promising frontier" (Ewert, 2020) in behavioral public policy as most applications in the past have treated behavioral measures as standalone solutions (Hansen, 2018; Sanders et al., 2018). Yet, behavioral measures were said to be more effective if taking into account the wider (policy) context (de Ridder et al., forthcoming). With this study we aim to take a step in that direction.

One of the main challenges to behavior change is the intention-behavior gap (Sheeran & Webb, 2016) where individuals fail to follow-up on their intentions. Assuming that canvassing affected intention but failed to bridge the intention behavior gap, we complemented canvassing using two nudges (Thaler & Sunstein, 2008). Nudges are light touch interventions that require little cognitive engagement from those targeted by the nudges (Thaler & Sunstein, 2008). This is because nudges tend to trigger automatic cognitive processes (e.g., biases and heuristics) in those targeted by the nudges bringing about predictable behavior change in a more subtle way. People are for example always more likely to select the default option, independently of its content (e.g., Jachimowicz et al., 2019; van Kleef et al., 2018). Nudges can be effective for mindless and subconscious behaviors, like littering (French, 2011). However, previous research shows that interventions directed at breaking unconscious behavior and making residents aware of and reflect on the challenges of their neighborhood (e.g., burglary) also have the potential to evoke long lasting behavior changes that benefit the community (Roach et al., 2020). In this study we therefore tried to stimulate both conscious and unconscious processes. The first nudge asked individuals reached by canvassing to show *commitment* to keeping the neighborhood clean by placing a sticker on their doorpost. Commitment nudges have been shown to harbor a large potential in evoking pro-environmental behaviors (e.g., Baca-Motes et al., 2012). Additionally, we expect that the canvassing itself will make people consciously think about the littering challenges in their neighborhood. For the second nudge visual *reminders* were employed that depicted the desired behavior and focused on strengthening a positive sense of *community identity* (i.e., by emphasizing group membership and shared responsibility for the neighborhood; Kolodko et al., 2016). Strengthening community identity has been shown to be effective for evoking pro-environmental behaviors before (Van Vugt, 2009). We assumed an appeal to community identity to be effective because the study area was characterized by high levels of social cohesion, which is associated by increased receptivity to community norms (Forrest & Kearns, 2001). This way, the nudges were integrated with the existing policy structure and city context rather than a standalone approach.

We compared both the standard canvassing policy and the same policy complemented with nudges to a control treatment, in which no actions directed at reducing illegal garbage disposals were carried out. We hypothesized that:

When comparing the pre-test with the post-test or follow-up, the number of days garbage is illegally disposed would be reduced after carrying out either the standard policy or the standard policy enriched with nudges, with a larger reduction for the nudging treatment.

## Materials and Methods

### *Environmental Context*

This study took place in a neighborhood (i.e., Oude Westen) close to the city center of Rotterdam. We made use of a convenience sample in which the control treatment (eight container locations) was

carried out in one area, the standard policy treatment (12 container locations) in another area, and the nudging treatment (10 containers and seven at follow-up due to road work) in a third area (Figure 4.1).

**Figure 4.1**

Map of the study area



Note: control treatment = right, standard treatment = center, nudging treatment = left, circles = container locations.

In Rotterdam, placing garbage outside garbage containers was illegal and resulted in a fine between €95 and €500 if detected. Public garbage containers were typically located underground with only a small part of the container being visible. To dispose garbage, residents used a container opening at hip height. Based on the amount of produced garbage, the frequency of emptying containers differed per city area. In the study area, containers were emptied when sensors inside the containers gave a digital signal that the container was almost full. Garbage outside the containers was collected daily, also during the experiment. For disposing items too big for the container openings (i.e., bulky waste) residents needed to arrange free individual pick-up or could bring their bulky waste to depot recycling.

The study area was densely populated (1,100 households) and characterized by high levels of ethnic diversity (71% of the residents had a migration background; compared to 49% on city level; Neighborhood Profile Rotterdam, 2016) and a relatively low mean income (65% of the residents were defined as receiving a “low” income; compared to 51% on city level). The vast majority of residents was between 15 and 65 years old (72%). Residents of the study area reported that illegal littering caused them substantial nuisance: 55% often experienced annoyance due to garbage placed

outside containers (compared to 48% on city level). However, different from other neighborhoods with serious littering problems, social cohesion was high: 62% of the residents indicated that they felt connected to the neighborhood (compared to 55% on city level). Treatment areas were all part of the same neighborhood, with comparable housing and population characteristics. No substantial differences between treatment areas were therefore expected. The combination of problem severity and high social cohesion made this neighborhood very suitable for an intervention targeting illegal littering by exploiting social commitments.

The municipality of Rotterdam took initiative in carrying out this experiment and consulted the Behavioral Insights Group Rotterdam (BIG'R) on the experimental design and procedure. BIG'R consists of municipality employees and behavioral scientists from Erasmus University Rotterdam who collaborate to improve public policy. BIG'R is thus comparable in its aim and activities to the well-known BIT UK (John, 2014).

**Figure 4.2**

Commitment stickers



Note: Commitment stickers on door(post) "I keep our street clean. Outdoors belongs to us all."

### ***Treatments***

The nudging treatment encompassed two components: the door-to-door canvassing and the placement of reminder boards close to containers. The canvassing was carried out during a 2 week's intervention period where, during 9 weekdays and one Saturday, five public information officers employed by the municipality reached 39% of households with maximum two attempts. Information officers were instructed to ask residents if they were familiar with the Rotterdam rules for garbage disposal, to explain them and inform households if necessary, and to provide households with a brochure summarizing rules and regulations for garbage disposal including some consequences of illegal garbage disposal. Importantly, only for the intervention treatment information officers also asked households to demonstrate commitment to keeping the neighborhood clean by placing a sticker

(Figure 4.2) on or near their door or doorpost. With 74%, most of the reached households complied. In total, 29% of households in the treatment thus received the full treatment.

From the second intervention week, information boards (Figure 4.3) were placed next to containers for the nudge treatment. These information boards remained in place at least until follow-up. Both the stickers and the boards emphasized shared responsibility for a clean neighborhood. Additionally, the board contained clear instructions for performing the desired behavior, both in written text and graphically. Due to practical reasons and municipality policy, the two nudges were thus integrated and overlapped in time making it impossible to evaluate them separately.

**Figure 4.3**

Reminder board next to containers



Note: Reminder board next to containers “Throw your garbage in the container. Together we keep our street clean. Outdoors belongs to us all.”

In the standard policy treatment, information officers conducted the same canvassing activities as in the intervention treatment. However, no commitment stickers and reminder boards were used. Note that for practical reasons the same information officers needed to conduct both canvassing treatments and could not be blinded to the different treatments. Public information officers reached 32% of households in this treatment. In the control area no actions regarding garbage disposal were carried out.

### ***Data Collection Procedure and Outcome Measures***

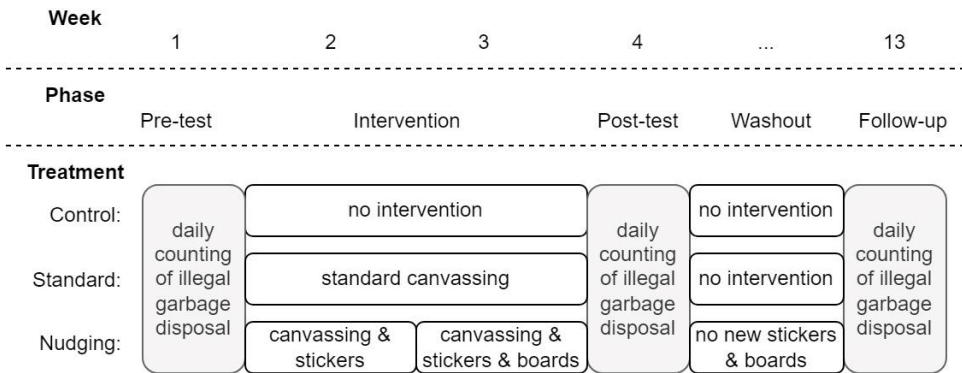
Garbage outside containers was measured during 1 week for each measurement. The pre-test measurement took place in week 1 (October 29 – November 3, 2019), the post-test measurement took place in week 4, and the follow-up measurement took place in week 13 (see Figure 4.4). This week was chosen for the follow-up because day length was comparable at that time to the post-test. Also, end of January was long enough after the holiday season, in which divergent garbage disposal could be expected (e.g., due to different working hours, and deviant production of garbage).

Only observational data on garbage disposals was collected and no personal data was registered. As a result, ethical review and approval as well as active consent from residents were not required for this study in accordance with local legislation and institutional requirements.

Illegal garbage disposals next to containers were recorded once a day by trained research assistants between 6:00 and 9:00 a.m. They registered if garbage was found outside each container (within a 5 m radius from the container), as well as the amount and type of garbage. Specifically, pictures were made, allowing researchers to check if new pieces of garbage had been placed or if it was remaining from previous days, i.e., garbage not picked up by the cleaning services of the municipality. Although the municipality tried to remove all displaced garbage daily early during the day, at some instances garbage was not picked up the same day. This was for example the case when it was hard to remove (e.g., a bucket chained to the container itself), or because of large items. At 11 instances (1.8% of all inspections) the same garbage was found a day later, and garbage registered the previous day was initially recorded as new illegally placed garbage. This was corrected in the data used for the analysis. Research assistants also checked and registered daily how the general treatment of containers was and if containers were full or hard to open. The order in which containers were checked varied each day. In the intervention area, research assistants also checked if the information boards were still in place and in good condition. If stickers were still visible and in place was not checked systematically, because information officers did not record which residents placed a sticker on their door (post). However, research assistants reported that stickers were still frequently present in the intervention area during follow-up.

**Figure 4.4**

Research planning per treatment



As an outcome measure, we selected the number of days garbage was displaced near a container rather than the amount of displaced garbage. This approach was chosen because it was hard to quantify littering *behavior* by number, kind, or weight of items for example. Ten empty bottles are hardly comparable to one piece of furniture but may result from the exact same behavior of a single resident person. Thus, it was unclear if multiple items were the result of one or more littering

instances from a single resident or multiple residents. Additionally, regardless of the number of items, additional cleaning needed to take place if any littering was detected. Using the number of days as an outcome variable was therefore also of high practical relevance. Therefore, the number of days that *new* placements were found was considered the main outcome variable.

### **Data Analysis**

Descriptive statistics from the pre-test were used to give an indication of baseline garbage disposals next to containers. Second, we compared the number of days per week on which garbage was illegally placed outside the containers between treatments: We used a mixed-design ANOVA with measurement time as a within-subjects variable, and treatment as a between-subjects variable. In case of significant main or interaction effects, repeated within-subjects contrasts for both treatment and time were applied to specify the results.

## **Results**

### **Baseline Garbage Disposal**

During the pre-test measurement, newly displaced garbage outside containers was found on average on 4.07 ( $SD = 1.82$ ) days a week across treatments. Most garbage fell within the category of bulky waste (45.6% of all 1,777 items), followed by paper (25.2%) and household waste (17.3%). In all three measurement periods, the highest proportion of containers with outside garbage was measured on Mondays (87%) and Sundays (55%), indicating that illegal garbage disposal was most common during the weekend.

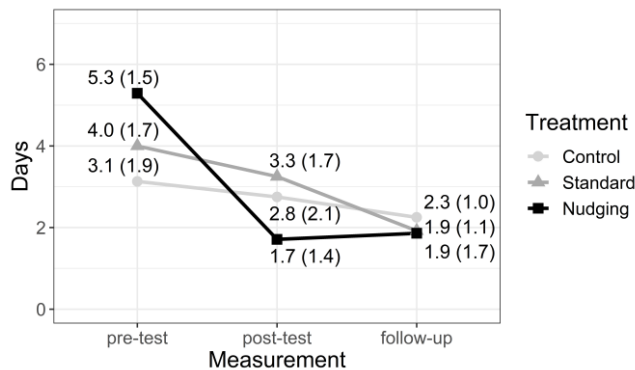
### **Effect of Treatment on Illegal Garbage Disposal**

For the analysis, only containers for which there were no missing data were included (i.e., three containers of the commitment treatment were excluded because they were not accessible at follow-up). The assumption of sphericity was met [*Mauchly's*  $W(2) = 0.859$ ,  $p = 0.173$ ]. There was no evidence for a main effect of the treatments on the number of days garbage was found outside containers [ $F(2, 24) = 0.22$ ,  $p = 0.803$ ], meaning that combined over the three measurement points the treatments were the same in the average number of days where illegal garbage was registered. However, the main effect of measurement time [ $F(2, 48) = 17.48$ ,  $p < 0.001$ ] was significant. Slopes differed between the pre-test and the post-test [ $F(2, 24) = 5.47$ ,  $p = 0.012$ ], but not between the post-test and follow-up [ $F(2, 24) = 2.27$ ,  $p = 0.126$ ], indicating that in general the treatments lead to a reduction in illegal garbage disposal directly after the intervention and remained stable until follow-up.

As expected, the interaction between measurement time and treatment was significant [ $F(4, 48) = 3.78$ ,  $p = 0.009$ ; Figure 4.5], indicating that garbage disposals developed differently across treatments.

**Figure 4.5**

Mean (SD) number of days new garbage was found outside containers per treatment and measurement time



Inspecting treatments individually across measurement times, no differences between time points were detected for the control treatment [ $F(2, 14) = 0.82, p = 0.460$ ]. Although a negative trend in the number of garbage disposals seemed to emerge, as expected, no significant decrease in illegal garbage disposals was found. However, the mean number of days that garbage was found outside containers differed between measurement times for the standard treatment [ $F(2, 22) = 10.52, p = 0.001$ ] and the nudging treatment [ $F(2, 12) = 10.93, p = 0.002$ ]: For the standard treatment this result was attributable to the difference between the post-test and follow-up [ $F(1, 11) = 1.94, p = 0.191$ ; effect size:  $d = -0.84$ ]. This means that no effect of the standard treatment was found directly after the intervention, while at follow-up a decrease was detectable. For the nudging treatment a decrease between pre-test and post-test ( $F(1, 6) = 0.05, p = 0.003$ ; effect size:  $d = 2.40$ ) was found, indicating that for this treatment results were observable right after the intervention and remained stable until follow-up.

When the mixed ANOVA is repeated including just the pre-test and the post-test, all container locations could be included in the analysis. This analysis confirmed the main results with a significant decrease of disposals found only for the nudging treatment [ $F(1, 9) = 33.92, p < 0.001$ ]. When comparing only pre-test and follow-up, a significant decrease in disposals was found for the nudging treatment [ $F(1, 6) = 21.13, p = 0.004$ ], as well as the standard canvassing treatment [ $F(1, 11) = 17.60, p = 0.001$ ], but not for the control treatment [ $F(1, 7) = 1.49, p = 0.262$ ].

## Discussion

We enriched the standard policy of the municipality of Rotterdam regarding illegal garbage disposals next to street containers (i.e., door-to-door canvassing) with a combination of nudges that complemented already existing municipality policy because this approach was expected to be most effective (de Ridder et al., forthcoming). Specifically, we employed a commitment-nudge (e.g., King et al., 2013) and reminders with clear, explicit, and graphic instructions for the desired behavior (Kolodko



et al., 2016). This approach was compared to the standard policy and a control treatment where no actions were taken to reduce illegal garbage disposal. At post-test, no effect of the standard policy or the control treatment was found. Adding nudges to the standard policy was highly effective at post-test, reducing the number of days garbage was illegally displaced with more than two-thirds and results remained stable until follow-up. This is an exceptionally large and lasting effect for a nudge intervention (Hummel & Maedche, 2019), with important consequences not just for municipality cleaning costs, but plausibly also for perceived neighborhood aesthetics (Roda et al., 2016) and perceived neighborhood satisfaction (Casakin & Reizer, 2017).

At follow-up, a decrease was also found in the area that received the standard policy and a decrease in illegal garbage disposals was found, resulting in comparable levels of illegal garbage disposal to the nudging treatment. For the nudging treatment and the control treatment, the effects remained stable at follow-up. The nudging treatment thus led to an immediate and lasting reduction in illegal garbage disposals whereas the standard policy led to a delayed and somewhat smaller reduction in illegal garbage disposals.

This delayed effectivity of the standard approach can be explained by the reach of the door-to-door canvassing campaign (32% of residents). Assuming that providing households with information brochures is an effective strategy to reduce illegal garbage disposals, still only those households that were reached could be expected to change their behavior directly after the intervention. Because no visible nudges were placed outside, other residents could not learn about the intervention other than by noticing a decrease in illegal garbage disposals. This decrease in garbage disposals could in turn signal a changing descriptive social norm (e.g., Cialdini, 2007; Sparkman & Walton, 2017), leading to less illegal displacements and littering by other residents (e.g., de Kort et al., 2008), just as visible garbage invites more littering behavior (Keizer et al., 2011). Research shows that prosocial behavior can be contagious, especially when social proximity is high (Dimant, 2019), as it was in the studied area. The time needed for such self-reinforcing cycle of positive behavior change to unfold could explain both the lack of results at post-test as well as positive outcomes at follow-up in the standard canvassing treatment. However, an alternative explanation for the delayed effectivity of the standard policy could be that residents from the standard policy area learned over time from the adjacent neighborhood in which (visible) nudges were applied. Residents from the standard treatment area could have noticed both stickers and information boards, as well as the changed descriptive social norm in the nudging area, resulting in less littering in their own area. In general, the physical proximity of the three different treatment areas means that our results may be affected by contamination effects. Future studies can be conducted in more dispersed treatment areas; however, this might increase differences between areas.

Like all research, this study has its limitations. First, because both nudges were tested together, it remains unclear which nudge (or their combination) was responsible for the reduction in illegal garbage disposals. This can, however, be investigated in future research. Second, characteristics of the study area have implications for the generalizability of our findings: A large proportion of neighborhood residents had a non-western immigrant background (60%, as compared to 37% on city

level; Neighborhood Profile Rotterdam, 2016). For a part of this group, language problems may have limited effectiveness of the intervention. Public administration officers involved in the canvassing campaign indeed indicated that they frequently encountered language barriers. Third, the study area was a neighborhood with relatively high levels of social cohesion. When conducted in neighborhoods with lower levels of social cohesion, the commitment and the community identity interventions may be less meaningful for the direct social environment and therefore less effective. Fourth, contextual factors that could not be influenced by the researchers (e.g., weather conditions) could have influenced dumping garbage disposal behavior. However, most likely, these treatments would have affected all areas equally since target neighborhoods were next to each other. Proximity of neighborhoods did, however, also come with a downside: We cannot rule out that effects found in the standard canvassing treatment were (partly) due to contamination effects from the adjacent intervention area. Passing regularly through a cleaner adjacent neighborhood may have strengthened the social norm and may thereby have reduced littering behavior. This effect might even have been strengthened by the canvassing itself, since this could have made the desired behavior more salient (Cialdini et al., 1990).

Lastly, between post-test and follow-up, three container locations for the nudging treatment were relocated outside of the research area, which could have influenced results because it complicated showing the desired behavior (i.e., disposing garbage in the containers). However, even under treatments possibly provoking illegal garbage disposals, because fewer containers were available, the effect of the nudging treatment remained stable.

Different from common nudging attempts to reduce littering that often involve enhanced visibility (e.g., colored bins, footsteps), our intervention was purposefully designed to complement existing measures enhancing their effectiveness at low cost. In addition, the employed nudges strengthened positive social norms that plausibly lead to more durable and robust effects than enhanced visibility (Lin et al., 2017). Yet, nudges in general have been criticized for being mere “fixes” that fail to challenge or change societal structures and patterns of behavior (Selinger & Whyte, 2012; Whitehead et al., 2017). Following this line of reasoning, our approach did not address the underlying problem of (large amounts of) garbage being produced. Moreover, nudges were said to embrace a narrow definition of autonomy allowing experts to paternalistically program behavior (e.g., Hausman & Welch, 2010; John & Stoker, 2019). In fact, although the nudges were transparent (i.e., their intention was obvious to residents) residents were most likely not entirely aware of their working mechanism and how their behavior was intended to be changed. Yet, it comes as a strength of this research that residents were unaware of the research (i.e., a natural experiment), increasing ecological validity.

## Conclusion

Illegal garbage disposals were a persistent and serious problem in Rotterdam, resulting in high cleaning costs and decreased satisfaction with the neighborhood among residents. Enriching the standard canvassing policy of the municipality (i.e., door-to-door canvassing) with *nudges* that emphasized community identity and shared responsibility, evoked commitment, and provided

reminders resulted in a two-third decrease of illegal garbage disposals when compared to the pre-test both at post-test and follow-up. This approach is thus highly promising in decreasing illegal garbage disposals to ultimately reduce cleaning costs, improve the aesthetical quality of urban areas, and reduce nuisance. In general, adding commitment strategies might be highly effective in improving canvassing at low cost. However, further testing with different neighborhoods is needed to judge the potential of this approach.



## Chapter 5

### No Pyro, No Party

**Social Factors, Deliberate Choices, and Shared Fan Culture Determine the Use of Illegal Fireworks in a Soccer Stadium**

This chapter is awaiting peer-review and has been published as:

Merkelbach, I., Dewies, M., Noordzij, G., & Denktas, S. (forthcoming). No pyro, no party: Social factors, deliberate choices, and shared fan culture determine the use of illegal fireworks in a soccer stadium [version 1; peer review: awaiting peer review]. *F1000Research*.

## Abstract

Lighting illegal fireworks inside soccer stadiums is a worldwide and persistent problem. Despite rules and regulations as well as rigorous enforcement, the use of illegal fireworks in football stadiums is increasing rather than decreasing. Little is yet known about the causes and predictors of this behavior, preventing the development of effective interventions or communication strategies. We therefore conducted a qualitative study, using semi-structured interviews with both supporters of a large Dutch soccer club who participated in lighting fireworks, and with professionals who dealt with illegal fireworks in daily practice. Semi-structured interviews were based on hooliganism literature as well as the COM-B model. We concluded that hooliganism and lighting illegal fireworks are distinctly different phenomena, although they share some underlying constructs. From a behavioral perspective, using the COM-B model as a framework, reflective motivation was identified as the strongest facilitator of lighting fireworks, which appeared to be an important part of supporter life and shared culture. Quick interventions that target automatic behavior, such as nudges, will probably thus not be successful in changing this behavior. Supporters suggest compromise between supporters and professionals as preferred future direction. Reported feelings of stigmatization and feeling unappreciated by professionals, could interfere with successful implementation of this direction. Professionals however contradict to have negative judgements of supporters. Building a bridge between supporters and professionals should be a first step towards a solution.

In many countries around the world, supporters lighting fireworks in soccer stadiums is considered a growing problem. This is also the case in The Netherlands where lightning fireworks is the most often observed punishable offense in soccer stadiums (Nohlen & van Harreveld, 2017) and where the number of incidents related to fireworks in soccer stadiums keeps increasing (e.g., Politie, 2018). Yet, lighting fireworks in soccer stadiums is illegal in The Netherlands as fireworks can legally be lit only around new year (article 2.1.1 of the Dutch Fireworks Decree). In addition, lighting fireworks is against regulations of the national (Royal Dutch Football Association; KNVB), the European (Union of European Football Associations; UEFA) and the global soccer associations (Fédération Internationale de Football Association: FIFA). Lighting fireworks concerns the safety of those inside the stadium as it is often difficult to keep the European recommended safety distance from fireworks of 1 to 15 meters (Guideline 2007/23/EG). Apart from potential direct damage or injury, lighting fireworks in stadiums was shown to cause a release of large amounts of nanoparticles, among which heavy metals, affecting health and the environment (Pirker et al., 2020). Containing the lighting of fireworks is thus important.

Efforts to contain the lighting of fireworks are ample and directed towards soccer clubs as well as supporters. The UEFA fines soccer clubs if a club's supporters light fireworks. In addition, the UEFA and the KNVB can require clubs to ban supporters from specific stadium boxes or the whole stadium for one or multiple matches. Individual supporters who are caught carrying or lighting fireworks inside the stadium or the stadium area receive fines from the KNVB and the Dutch government adding up to approximately 900 euros. In addition to these fines, those supporters receive a stadium ban of minimum six months. To detect fireworks offenders, soccer clubs frisk supporters at the entrances to the stadium area and surveillance is used, namely cameras, stewards, and the police. In short, fireworks in soccer stadiums affects many people, involves high costs, and demands much effort from clubs, stewards, police, and prosecution.

The high number of fireworks related incidents shows that the behavior continues and that current repressive efforts are not effective (enough) (Auditteam Voetbal en Veiligheid, 2017). Increasing restrictive measures even further has multiple downsides because it may not be effective, it is expensive, and it is likely to make the atmosphere in the stadium less pleasant and welcoming especially for the majority of complying supporters. For instance, when waiting times at stadiums entrances are increased because of intensified frisking. Behavior change strategies grounded in a better understanding of the behavior of lighting fireworks are currently seldomly used, but may be effective, less restrictive, and cheaper. Therefore, we took a behavioral approach interviewing soccer supporters and professionals to understand the use of fireworks in the stadium. Specifically, we focused on social, individual, and contextual factors that could explain the behavior. Behavioral explanations are crucial for identifying effective behavior change strategies (e.g., Howlett et al., 2020; Michie et al., 2011).

This research was initiated when the municipality of Rotterdam approached the Behavioral Insights Group Rotterdam (BIG'R) in aiming to design strategies to contain the use of fireworks in the stadium of Rotterdam's largest soccer club, Feyenoord. Like other behavioral insights teams around

the world (Afif et al., 2018), BIG'R enables institutionalized collaborations between behavioral scientists and public servants to pioneer the application of behavioral insights for public policy (John, 2014). Feyenoord is one of the major soccer clubs of the Netherlands with on average about 45.000 supporters visiting home matches in the Feyenoord stadium (KNVB Expertise, 2015). Feyenoord is renowned for the extensive use of fireworks by the club's supporters (Hellinga, 2019).

To the best of our knowledge, (behavioral) scientific literature on lighting fireworks in stadiums is limited, leaving the phenomenon largely unexplained. When lighting fireworks, supporters however necessarily make planned efforts to not get caught and they risk adverse consequences (stadium bans, fines). This suggests that they perceive this behavior as important or valuable. Indeed, Brechbühl and colleagues (2017) report that lighting fireworks is seen as important and likely part of a shared culture and identity for fanatical soccer supporters. However, the authors do not analyze the behavior of illegally lighting fireworks in depth since not fireworks, but markers of escalation are the prime focus of their research. The notion that fireworks are important to substantial groups of supporters is supported by the existence of several websites (e.g., In de Hekken), songs (e.g., "No Pyro No Party"), blogs, and YouTube profiles that are created and managed by supporters which discuss the subject in a positive manner.

Fireworks offenders may (in part) be the same social group as hooligans because fireworks tend to be lit by fanatic supporters and in stadium boxes that are also popular among hooligans. Based on a survey, Scholz and Hurych (2018) states that the majority of soccer hooligans indeed also engages in lighting fireworks, suggesting at least a partial overlap between supporters who light fireworks and hooligans. In the absence of a body of literature specifically on lighting fireworks by supporters, what is known about supporters and hooliganism is used to inform and contextualize our research. Soccer hooliganism, defined as violent and destructive behavior (Budim, 2018), is often explained by group identity and a shared culture (Zimniak, 2020) in which the *ingroup* is contrasted with the *outgroup* (Ward, 2002). Within the ingroup, strong social ties, social identity within the group, and the use of symbolism are of great importance (Dionísio et al., 2008). Spaaij (2008) identifies six fundamental features of the hooligan identity: excitement and pleasurable emotional arousal, hard masculinity, territorial identifications, individual and collective management of reputation, a sense of solidarity and belonging, and representations of sovereignty and autonomy. Therefore, ingroup and identity features are candidates for explaining the use of fireworks.

Scholz and Hurych (2018) however categorized less than 5% of fanatical supporters (i.e., supporters visiting the most fanatical box of boxes in the stadium) as hooligans, and it seems that for most coordinated fireworks events more than 5% of fanatical supporters is involved. This suggests that at least part of all fireworks lit inside the stadium cannot be attributed to hooligans. How large the overlap is between supporters who light fireworks and hooligans, is currently unknown. Hooligans are known to be supported in their behaviors by a group of other supporters who have a positive attitude towards hooligans, so called *hoolifans* (Rookwood & Pearson, 2012). Supporters who light fireworks may be supported in a similar way by other supporters who encourage and reward that behavior. For instance, fanatical female soccer supporters from Czech Republic indicate to almost never light



fireworks but almost exclusively regard it as a positive act and sometimes even assist (Scholz, 2020). They may not do it because fanatical fandom in soccer stadium is still associated with traditional masculinity (Kossakowski et al., 2020).

In the current study, we chose a qualitative design and took an interview approach. Explorative in-depth interviews are suitable for collecting information on complex (and understudied) phenomena (Sofaer, 1999). In order to gain understanding of both the perceived barriers as well as motivations for and facilitators of the behavior, interviews were based on elements described in hooliganism literature and the COM-B model (Michie et al., 2011). This general behavioral model assumes that capability (physical and psychological), opportunity (social and physical), and motivation (automatic and reflective) are basic conditions for the occurrence of all human behavior. We chose this model because we considered it broad enough to accommodate the elements from the hooliganism literature (such as social environment and motivation) and because it could guide the exploration of other personal, social, and contextual factors affecting the behavior. Additionally, we investigated possible solutions and visions for the future. We interviewed supporters who frequently lit fireworks at the time of the interview or sometime in the past. Some had recently received a stadium ban for lighting fireworks. As a frame of reference, we also interviewed professionals who dealt with fireworks in the stadium as part of their work, e.g., employees of Feyenoord and police officers. Permission to conduct this research was obtained from the ethics review committee DPECS, ESSB, of the Erasmus University Rotterdam (19-045).

## Methods

### *Participants*

#### *Supporters*

16 fanatical supporters of Feyenoord participated in this study (i.e., all participating supporters had, or recently had had a season ticket). Data collection stopped at this point because the first author conducting all interviews felt saturation was reached because no new information came up in the last interviews. Although sample size is small, we believe we gained a complete picture of the behavior, because the studied behavior was very specific, and the target group relatively homogenous. We used purposive sampling and interviewed both supporters who were lighting fireworks at the time of the interview and supporters who had stopped doing so (both making up half of our sample). All supporters had lit fireworks during matches and/or Feyenoord-related events such as training sessions. Additionally, we interviewed supporters who had and had not been caught (half of the supporters currently lighting fireworks were caught as were a third of the supporters not lighting fireworks any longer). Participants were recruited a) through the official supporters association of Feyenoord or b) were known by the club because they carried out chores in order to shorten a recently received stadium ban related to fireworks.

All participants were men with a Dutch background. Participants were between 20 to 49 years of age, with a mean of 30 years. Several supporters showed their fandom through their appearance, for

example by wearing Feyenoord clothing and tattoos. Some of the supporters who were interviewed via video calling also used Feyenoord related items for their interior design.

### *Professionals*

10 professionals participated in this study to offer an expert frame of reference. We again made use of purposive sampling trying to include at least one representative of all major institutions dealing with the issue. Participants worked at the municipality of Rotterdam ( $n = 1$ ), soccer club Feyenoord ( $n = 3$ ), the supporters association ( $n = 1$ ), the police and justice department ( $n = 4$ ), or a professional fireworks company that had conducted several legal fireworks shows in the Feyenoord stadium ( $n = 1$ ). Data collection stopped after these 10 interviews because all major professional groups dealing with the issue had participated, and information provided by professionals within the same institution was highly similar. Professionals at higher management levels (e.g., KNVB or UEFA) could not be reached.

### **Procedure**

#### *Preparation and Start-Up Phase*

The municipality of Rotterdam contacted BIG'R to contain the use of illegal fireworks in the Feyenoord soccer stadium. Several exploratory conversations with the municipality, the club, and the police and justice department took place. Due to the complexity and persistence of the problem and the lack of scientific literature on the subject, it was decided to carry out a qualitative study aiming to increase understanding of the behavior.

#### *Recruitment, Selection, and Requirements*

Professionals were recruited via the professional network of the contact person of the municipality of Rotterdam. In two instances colleagues of these professionals were interviewed after a professional had suggested to interview them because of their relevant knowledge.

Via professional contacts of Feyenoord or the Feyenoord supporter association, supporters received an information letter about this research. They could then give written consent to be approached by the research team. Supporters were approached in the name of research (rather than in the name of the municipality) in order to ensure they felt comfortable to take part in the study. Supporters were assured they would not experience any negative consequences if they participated in this research and talked openly about their behavior. They were informed that their participation was an opportunity to have a voice in the public debate as the outcomes of this research would be shared with various institutions in order to facilitate finding solutions. Participants who agreed to participate gave written consent before the interview. For their participation, supporters received a Feyenoord gift card worth 15 euros. Professionals did not receive any reimbursement for their participation.

#### *Interview*

Because the topic was sensitive, we decided to conduct one-on-one interviews. Interviews with professionals took place at their work locations and interviews with supporters took place in the

Feyenoord stadium, a room offered by the Feyenoord supporter association, or via video calling due to COVID-19 measures. Supporters generally came across as motivated and eloquent. On average, interviews lasted about 50 minutes (range 25 to 90 minutes). The interviews were audio recorded and upon request participants could receive the audio recording or a transcribed version of their interview. However, none of the participants used this opportunity. After the interview, supporters were asked via email to anonymously answer a short online questionnaire evaluating the interview experience. From 12 who could be contacted via email, 10 filled in the survey (83%). In general, supporters experienced the interview as pleasant, and as an opportunity to share their opinion (see Table 5.1).

**Table 5.1**

Evaluation of interview experience by supporters

Item	Mean (range)
I was able to give a complete picture of my opinion during the interview.	4,7 (4 – 5)
During the interview, I felt free to speak openly and honestly.	4,8 (4 – 5)
I felt that I was being listened to during the interview.	4,8 (4 – 5)
I have experienced the interview as pleasant.	4,8 (4 – 5)
Relevant questions were asked during the interview.	4,5 (4 – 5)

Note: 1 = totally disagree, 5 = totally agree.

## **Materials**

### *Interview Guide*

All participants were asked the same set of questions (interview guides for supporters and professionals can be found in the online supplementary material at <https://doi.org/10.5281/zenodo.4835766>) which could be complemented with questions based on the answers of participants (i.e., semi-structured interview). Questions were partly based on hooliganism literature, and partly on the COM-B model (Michie et al., 2011).

### *Evaluation of the Interview Experience*

To be able to assess to what extent supporters felt they had been able to speak freely during the interview, they filled out a five-item questionnaire (see Table 5.1). Professionals did not fill out an evaluation because socially desirable answering or reluctance to provide answers was not expected to be an issue in this group since they did not report on their own illegal behavior.

## **Coding and Analysis**

All interviews were literally transcribed and coded by two researchers, using thematic analysis (Braun & Clarke, 2006). Codes were obtained through inductive coding in Atlas.ti 8.4 and grouped into categories. If multiple codes were applicable to the same piece of text, codes sometimes could overlap. Differences of opinion between coders were solved through discussion and consensus. Interviews of supporters were double coded, meaning that each quote received both a category and a

subcategory (i.e., code) label. Because professional interviews were not considered the prime focus of the current study and are used as a frame of reference, only categories were coded here.

## Results

During analysis we identified five categories with each code belonging to one of them. The categories were: Ingroup/Feyenoord culture vs. Outgroup (27 codes), Context (18 codes), Motivation for lightening fireworks (13 codes), Appreciation of measures and perceived barriers (21 codes), and Solutions and future perspectives (8 codes). See the online supplementary material (<https://doi.org/10.5281/zenodo.4835766>) for a complete overview of supporter codes and how often they appeared. Based on the five main categories, we present our findings. Professionals' opinions are also shortly discussed, offering a frame of reference.

### ***Ingroup/Feyenoord Culture vs. Outgroup***

All supporters report that they have been supporting Feyenoord from early age onwards. Most of them came in contact via their father, some via peers. Being a Feyenoord supporter is an important part of the own identity, in which not soccer but the social ties are appreciated as most important. For example, one of the supporters stated clearly that his social life is determined by connections with other supporters:

*"It really is the binding factor in my social life. All my friends have a season pass, it has always been the case that my entire life and social life is about soccer. In this phase of my life, we see each other at matches and at birthdays of our children. But back in the day, we did everything together. On Saturdays, Sundays, and by the time it was Thursday it was time to hit the pub again."*

Supporters describe a sharp dividing line between ingroup and outgroup. The ingroup is experienced as fanatic, loyal, and committed. Within this group, there is an established hierarchy, social control, and consideration for each other. Ingroup members participate in creating the atmosphere via fireworks, singing and cheering, or other actions. They have at least a season ticket to certain boxes, are seen as fanatical, and share club culture and social ties. Outsiders on the other hand are less important than ingroup members, however with varying appreciation for different kinds of outsiders. Other, less fanatical supporters who also come to the stadium are seen as relatively positive.

Supporters also believe that these outgroup supporters evaluate both the fanatical group as well as the fireworks actions as positive. Generally, fanatical supporters report not really thinking about or not having an opinion about people who watch games on television. The media, the KNVB, and most of the time the club (and its board) are negatively appreciated.

Professionals describe Feyenoord as a working-class club and believe that fireworks are seen by the supporters as part of the soccer experience. Some of them mention that fireworks are part of soccer culture.

Most supporters have a strong historical awareness and a sense of shared culture. Even the youngest supporters show a sense of nostalgia. Supporters believe that the supporter experience is currently restricted by measures and regulation, making the current experience less attractive. One supporter for example stated that people are no longer as enthusiastic about visiting matches as when fireworks were still allowed:

*“Back in the day, we were allowed to light fireworks, there even was a special box for it. And when the team trained at the public training facility outside, there was a crowd of 20.000 to 30.000 people, it was just one big conflagration. And they’ve killed that. We can’t do fireworks, we can’t do anything, and people just stopped coming.”*

Relatedly, a large part of supporters state that fanatical supporters are generally stigmatized, especially by people in positions of power. This aspect was often brought up by the supporters rather than being directly linked to the prepared interview questions. Supporters experience that they are in general unjustly portrayed negatively and valued by the media, Feyenoord, and the KNVB. They report that these institutions describe them overly negative with a heavy focus on small negative incidents. Additionally, supporters feel they are being actively thwarted. Supporters suspect that the club and the KNVB do not appreciate fanatical supporters and wish to ban them from the stadium. They are believed to employ unjust measures to reach their goal (such as handing out stadium bans for very small offences).

*“I really get the feeling they want to ban us; we are chased out of our stadium.”*

Professionals do however not speak negatively or in a stigmatizing manner about supporters. They characterize supporters who light fireworks as fanatical and highly involved in the club, both in spirit and action. They are seen as clearly different and separate from other supporters. Sometimes professionals use appreciative phrases like “real supporters” to describe the fanatical subgroup.

### **Context**

The context category describes the context in which fireworks were lit. This concerns the physical context (e.g., certain boxes), but also other more socially determined contextual aspects, such as unwritten rules. When it comes to social economic status and educational level, all supporters report a varied group of offenders. Generally, offenders are described as men between 16 and 40 years. Lighting fireworks generally happens coordinated and together with other supporters. Most often, coordinators of firework actions are experienced group members who typically do not light fireworks inside the stadium themselves. Lighting generally happens by younger members of the group.

Supporters chose to collectively light fireworks because of the visual effects of coordinated group/mass actions, but also to reduce chances of getting caught. Supporters indicate that they take their direct environment into consideration when lighting fireworks. They scan their surroundings, try not to hurt anyone, and take the presence of children and elderly into account. Additionally, a set of consistent unwritten rules are in place, such as “no throwing of fireworks”. Social control makes sure that these rules are generally respected and followed. The following statement illustrates that the lighting of fireworks is not used to cause any annoyance or inconvenience:

*“I don’t want other people to be bothered by my behavior. If someone close to me can’t stand the smoke, I would not do it, or at least I would give a warning.”*

Supporters mention that social pressure or ingroup reputation could stimulate young members of the fanatical group to feel pressure to light fireworks. However, none of the supporters indicate that they themselves ever felt pressure. Own responsibility and choice are important and are frequently mentioned.

Supporters describe that fireworks should not be lit during each and every game but should be saved for special occasions. Especially evening matches (because of the stronger visual effect), matches against larger clubs or rivals, and matches of high stakes are mentioned. Fireworks are also used for other occasions with a large emotional component, such as weddings, funerals, and memorials of supporters. Fireworks thus seem to be a broad and important aspect of supporter culture, used during many different momentous situations and contexts, as is illustrated in the following comment:

*“I have been at a wedding of soccer friends where fireworks were lit, that really adds something extra. Or at funerals, when someone has died. The entire Erasmus bridge [iconic bridge in Rotterdam] lit up with torches, that really gives you goosebumps.”*

Supporters do not perceive fireworks as unsafe. They feel fireworks are safe because supporters take each other into account, experience, knowledge, and knowhow of offenders. Additionally, supporters report to never or hardly ever experience any fireworks related incidents or injuries.

Supporters make a clear distinction between smoke pots and torches, and bangers. The use of smoke pots and torches is broadly supported, while using bangers inside the stadium is broadly condemned. Some supporters even feel that supporters who do light bangers inside the stadium should be punished more severely than is currently happening. Supporters also mention that the current restrictions and regulations make lighting fireworks less safe. For offenders to not get caught, they wear face covering clothing and pass burning torches. Because fireworks are forbidden, no safety measures (such as buckets of sand) can be implemented in the stadium.

*“You know, part of the danger is really only caused by forbidding it.”*

Professionals describe the group of offenders as diverse, with a variety of backgrounds and educational levels. Professionals also see lighting fireworks mostly as a group activity and seem to be aware of the different methods supporters use to prevent identification and prosecution. They describe that sometimes supporters light fireworks individually ("lone wolves"). However, these offenders are caught easily with the major problem being coordinated group lightnings of fireworks or repeated offenders. Professionals agree that important matches are more likely to evoke firework actions. Moreover, they admit to sometimes being surprised by fireworks actions. Some professionals have the idea that supporters would want as many fireworks in the stadium as possible. While almost all professionals feel that supporters underestimate the risks that lighting fireworks inside the stadium encompasses, they do feel that supporters try to act safely and responsibly.

Some professionals do state that indeed the number of incidents is very low, but that potential incidents could turn out to have severe consequences. None of the professionals believes supporters to be indifferent, or to cause dangerous situations on purpose. The differentiation between bangers on the one hand and smoke bombs and torches on the other that is frequently made by supporters, is barely discussed by professionals. They seem to perceive fireworks as one category of products that represents danger to visitors of the stadium.

### ***Motivation for Lightning Fireworks***

All supporters unanimously indicate that enhancement of the atmosphere is the primary reason for lighting fireworks. Supporters feel fireworks to contribute to a positive and fanatical atmosphere, also whipping up less fanatical supporters to sing and cheer along. Through this enhancement of the atmosphere, fireworks are believed to have a positive effect on the performance of Feyenoord players and potentially, but less importantly, an intimidating effect on the opponent's players, as is described by the following supporter:

*"The moment these torches start burning, you immediately start hearing reactions in the stadium. They become enthusiastic; they start to make a noise. Like three years ago, we played Napoli. During the entire game, there weren't any torches, and we were on a draw, ready to be eliminated [from the competition]. So, we decided, five minutes before the match would end, to support our team one last time. So, in the 85<sup>th</sup> minute, we lit a sea of torches. Everyone reacted positively, the entire crowd. Everyone stood up, started singing and cheering, and eventually, in the last minute, we scored. The former trainer admitted, it was us that gave the final push."*

Offenders also enjoy lighting fireworks in the stadium. Additionally, reputation of the fanatical supporters group and competition with other supporter groups is named as an important motivation for lighting fireworks. When it comes to fireworks, interviewed supporters believe their reputation to be the best in The Netherlands. Living up to this reputation is perceived as very important, which is attempted

by organizing the most beautiful, biggest, most daring, and best coordinated fireworks actions. One supporter described it as:

*“When it comes to fireworks in the stadium, Feyenoord is the best of the Netherlands. Of course, you see it everywhere, but we are the best. Head and shoulders above.”*

Fireworks actions are shared on social media and several websites with both the own ingroup as well as with fanatical supporters of other teams. Re-watching a successful fireworks action gives a sense of pride and togetherness. Competition between different supporter groups is described as positive and almost friendly. Supporters indicate they welcome and applaud well-executed fireworks actions by other supporter groups. The following statement makes clear that no hostility but competition between supporter groups is experienced:

*“It is all very positive. Rivalry has always been associated with beating the living hell out of each other, but when it comes to fireworks, that is not the case at all. It is a positive rivalry, we are trying to outdo each other in a friendly manner, on the one hand with singing, on the other hand with fireworks.”*

Lastly, also some rebellion was named as a reason for lighting fireworks. Some supporters feel that the fact that lighting fireworks is forbidden, makes doing so even more thrilling and challenging. Supporters also indicate that harsh execution of measures directed at containing fireworks or introducing new measures, increases motivation to light more fireworks. This way, supporters try to show that they will not be controlled by any authorities, but they are the ones in charge of shaping their own supporter culture.

Entertaining other supporters, protest, peer pressure, and intimidation of the opponent are not seen as important motivations to light fireworks. Just like supporters, professionals also indicate they believe enhancing the atmosphere and revelry are the most important motivations for lighting fireworks. Most professionals indicate that they can understand these reasons and can empathize with the supporters.

Professionals are very positive about other supporter driven actions, like large banners or flags. They know that supporters believe that lighting fireworks will enhance the performance of players. Professionals also believe that intimidating the opponent is an important motivation for supporters. Most professionals state that they do not believe fireworks to affect players' performances. Professionals also mention competition with other supporter groups and personal reputation within the ingroup as reasons for lighting fireworks. While supporters do not mention peer pressure as an important reason for lighting fireworks, professionals mention it often.

Professionals indicate that there is a strict hierarchy in which young group member have to prove themselves to established members by (among other things) lighting fireworks.



### ***Appreciation of Measures and Perceived Barriers***

Supporters who have stopped fireworks give reaching a certain age or stage of life as most important reason for stopping. They also mention that the use of fireworks is not matching work or family life. However, in general supporters indicate that they experience few real barriers when it comes to lighting fireworks inside the stadium. Measures are perceived as ineffective, especially because the chances of getting caught are slim. If chances of getting caught were higher, a stadium ban would by far be the greatest barrier. Not being able to attend matches while peers can, is or would be experienced as very aversive. One supporter describes a stadium ban as follows:

*“You don’t even want to think about it, every game you are forced to watch on TV, you’re reminded you’re not in the stadium.”*

Supporters perceive the measures as exaggerated and patronizing. The fact that fireworks are not seen as causing any harm, plays an important role and receiving two fines (one from the KNVB and one from the government) is perceived as unfair. Most resistance is felt towards collective punishment. Especially the collective stadium ban is seen as out of proportion. Such punishments, which are perceived as unfair are said to even evoke more lighting of fireworks as an act of protest, instead of preventing the use of fireworks. Other measures are also perceived as counterproductive. Supporters state that because there are measures in place, lighting fireworks becomes less safe, and that less fireworks would hurt the achievements and reputation of the club, and thus also the club’s earnings.

Professionals indicate that measures are generally not effective (enough). The stadium ban is perceived as potentially most effective since it is seen as the most adverse. However, because of low chances of getting caught, this measure is also perceived as ineffective. Professionals thus share the opinion of supporters on this matter. Some professionals indicate that some measures are not effective at all (like frisking or the occasional use of sniffer dogs) but are implemented to showcase efforts to the KNVB and other organizations.

### ***Future Perspectives and Solutions***

Supporters unanimously indicate that fireworks make up an essential part of the supporter culture and can never be replaced by other products. While fireworks organized by the club are appreciated, this is not seen as a replacement for supporter driven fireworks. Real fireworks should be organized by supporters and should be spontaneous, rough, and real.

*“Nothing is as beautiful as an action organized by supporters, because it produces a certain spontaneity, intensity, and effectiveness.”*

As possible solutions, supporters suggest a box in which fireworks are allowed. In this box, safety measures could be installed, such as extinguishing facilities. Another solution frequently

mentioned is a ticket premium used to pay fines Feyenoord receives from the KNVB and UEFA. All supporters mention that they are willing to collaborate, compromise, and make agreements with the club for preparing fireworks actions together. However, both parties should be willing to compromise, as illustrated by the following quote:

*“Have supporters tell their story, what they find important. Now they’re just ruling from an ivory tower. They’re afraid to get among the people. But if they would, they would know what happens inside the stadium.”*

Professionals suggest an even broader range of solutions. They, for example, suggest more replacements products, like confetti and laser shows. Some professionals mention that an outspoken disapproval of fireworks from other supporters or players might result in less fireworks actions. However, those professionals who communicate with fanatical supporters more often feel otherwise. Some professionals advocate stricter measures (like prohibiting wearing hoodies inside the stadium, or conducting a house search in case of a caught offender).

Most professionals, however, state that they would prefer to see forms of legal and organized fireworks inside the stadium. Almost all professionals say that they would prefer to work together with supporters, would like to start a conversation, and are willing to aim for some form of compromise.

## Discussion

The illegal lighting of fireworks by supporters inside soccer stadiums is an understudied and a therefore little understood phenomenon. To gain insights in this behavior and potential solutions, we conducted an interview study with both supporters showing this behavior and professionals who deal with this behavior.

We chose the COM-B model (Michie et al., 2011) to accommodate our findings (see the online supplementary material at <https://doi.org/10.5281/zenodo.4835766>). Starting with capability, supporters are convinced that they have the physical as well as psychological means to use fireworks safely and they are mostly unaware of any incidents. When it comes to physical opportunities, supporters take their environment into account to avoid hurting others. However, the absence of any physical safety measures makes lightning fireworks less safe. Generally, supporters see no replacement for supporters-led fireworks, which is associated with specific stadium boxes and matches. Social opportunity aspects seem to play a major role for the use of fireworks. Relevant ingroup members often light fireworks as well, and generally other supporters are believed to welcome the use of fireworks. The social context also influences when (during which matches, at which moment) and which fireworks are used via unwritten rules and group hierarchies. An important aspect is that supporters with their firework actions stand in competition for reputation with supporters from other clubs. Social contexts outside the stadium (namely one's family and work requirements) may discourage individual supporters from the use of fireworks. Turning to motivation, it is obvious that lightning fireworks is strongly rooted in reflective motivation. One does not simply walk into the stadium

with a torch, somewhat coincidentally. Rather, supporters prepare and conduct their firework actions deliberately and in groups, often coordinating their behavior via social media, as becomes also clear from the quote by one of the supporters indicating that shortly before the end of the match against Napoli a great number of Feyenoord supporters lit fireworks together to support the team. In addition, they are required to take deliberate measures to avoid getting caught and receiving a stadium ban that is perceived as dreadful. In fact, lighting fireworks is considered one's individual choice. Supporters do not report vast amounts of peer pressure. Sometimes using fireworks is a collective act of protest or revelry against authorities or the club and their decisions. While lighting fireworks has a strong component of automatic motivations as it possesses strong positive qualities in terms of emotions (e.g., collective cheering) and senses (e.g., appealing images), lighting fireworks seems mostly a deliberate and conscious choice. It thus follows that simple, one-off behavior change techniques (e.g., nudging, communication campaigns) targeting individual supporters are unlikely to contain the use of fireworks and that more sustainable attempts targeting social groups (e.g., culture change) are more promising.

Indeed, the behavior should not just be considered an individual's behavior that can be understood from just a psychological or behavioral perspective but should also be contextualized within its social and cultural context. We do this by using insights from literature on hooliganism. When the six fundamental features of the shared hooligan identity as defined by Spaaij (2008) are considered, the identity features of fireworks offenders seem to be largely comparable. Excitement and pleasurable emotional arousal (1) are often reported as important motivations for lighting fireworks, as is the acquirement of a positive reputation, both individually and on a group level (2). When describing the *ingroup* (Ward, 2002), supporters often refer to strong social ties (3) and the need for being able to be their "own boss" or to do as they please in "their" stadium (4). However, although offenders are (almost) exclusively male, masculinity (5) is not reported as an important factor in explaining the illegal lighting of fireworks. Also, territorial identification (6), the felt need to defend the stadium and surrounding grounds and invade rival's territory (van Ham et al., forthcoming) is important to hooligans, but is not seen as important to supporters who light fireworks. Possibly because supporters of rival clubs who light fireworks are not perceived as hostile but as comparable and sometimes even respected groups, being part of the same culture and community. Rivalry between these groups is described as almost amicable. The aggressive component central to hooliganism (Budim, 2018), does not appear to play any role in explaining the illegal lighting of fireworks; supporters describe it as a positive behavior meant to enhance the atmosphere inside the stadium and to boost the performance of their team, making this behavior distinctly different. In short, it may be hooligans who light fireworks, but lighting fireworks does not make a supporter a hooligan. A conclusion that was confirmed by the views of professionals

Lighting fireworks can thus not be seen as just another aspect of hooligan culture and it clearly represents substantial cultural value, being a distinct behavior that is shared, valued, and learned by a specific group of people (Birukou et al., 2013). Within the *ingroup*, fireworks have substantial symbolic value, as illustrated by their use not only during important matches but also during important life

events such as weddings, funerals, and memorials. This symbolic appreciation of fireworks is not unique to soccer culture, but can be found in many cultures worldwide, often signaling new beginnings or festivities (e.g., Lynn, 2006). Supporters describe the lighting of fireworks not as a hobby or pastime, but rather as shared cultural heritage, eliciting a sense of nostalgia and shared identity, that is often even passed on within families, as is often the case with fandom and soccer passion in general (e.g., Gil, 2002). Appreciating fireworks as part of the cultural heritage of a distinct and close cultural group, could explain the importance supporters assign to this behavior. Most professionals do not describe the use of fireworks by supporters from this perspective. Most mention the cultural context and value of fireworks only briefly. This could indicate an underestimation of the experienced value of fireworks by supporters.

This gap between the appreciation of the (cultural) importance of fireworks by supporters and professionals could signal a lack of mutual understanding. These differences in perception, mutual incomprehension, and lack of perspective taking could be the breeding ground for the stigmatization supporters frequently report (Chung & Slater, 2013). Observational studies have shown that enforcement by authorities that is seen as too strict or unjust, could lead to more aggressive or unwanted behavior (e.g., Stott & Adang, 2009). This could also be the case when it comes to fireworks, since supporters often reported that they felt that professionals did not understand or appreciate them, consciously tried to thwart them, or even wanted or tried to ban them from the stadium all together. Supporters do thus not feel that they are treated properly and justly and, in line with procedural justice theory (Nagin & Telep, 2017), are thus not motivated to comply with laws and regulations.

However, professionals do not hold as negative views of the supporters as supporters themselves believe they do. Professionals do generally not approve of illegally lighting fireworks, but often express appreciation of the positive intentions motivating it, and many of them praise the visual qualities of fireworks actions. Professionals mostly do not oppose the *spirit* of the fireworks, but point to practical barriers such as fines and safety concerns. Also, when professionals talk about the group of fanatical supporters in general, they are mostly described positively. Previous research into the willingness to engage with hooligans suggests that the stimulation of mutual perspective taking can result in more willingness to establish and maintain contact (Wang et al., 2014).

Bridging the perceived opposition between supporters and professionals could be a first step in reaching a solution acceptable to both parties. Both professionals and supporters indicate that banning fireworks from the stadium entirely is not possible, nor desirable. They question thus, although only implicitly, the current legal situation that is prohibiting all fireworks. Both groups mention dialogue, compromise, and mutual agreements as preferred solutions. Also, alternative ways of lighting fireworks inside the stadium are proposed by both groups; professionals mostly suggest lighting fireworks in collaboration with professional companies, while supporters would appreciate actively contributing to the fireworks, for example in a fireworks box. When mutual trust and understanding is established, reaching consensus does not seem impossible, however current

national and international legislation and regulations could stand in the way of solutions preferred locally.

### ***Strengths and Limitations***

The current study is unique because it was able to recruit a normally hard to reach group (i.e., fanatical soccer supporters) to participate in research, asking them about a sensitive topic (i.e., illegal behavior). However, supporters unanimously indicated to feel safe to discuss this topic and their views matched the images painted by professionals, suggesting supporters provided reliable information. The combination of supporter interviews and interviews with professionals dealing with the (illegal) fireworks in the stadium, offered a broad and presumably valid picture of the behavior at hand. Lastly, the current study explored an understudied, though important, topic.

Unavoidably, the current study also has certain limitations. Most supporters were recruited via the Feyenoord supporter association, an organization which is known to endorse the (safe) use of fireworks inside the stadium, and which could thus be inclined to get across a (too) positive image of supporters who light fireworks. Possibly, this led to recruitment of the most eloquent and nuanced supporters, leaving out those who light fireworks with aggressive intentions. However, supporters who were recruited via Feyenoord, officially opposed to the use of fireworks by supporters, did not contradict these supporters but painted a similar picture.

Additionally, it is not clear to what extent the current findings are generalizable to other clubs or countries. However, research suggests that ultra (highly fanatical supporters) culture is largely comparable and even interconnected across Europe (Kennedy, 2013), while also the use of pyrotechnics is widespread, and described as a big, and often growing, problem all across Europe (Standing Committee of the European Convention on Spectator Violence, 2011). Testimonials of a fanatical supporter visiting different clubs and matches across the world to experience fan culture, including the use of fireworks, also suggest shared culture and similar motivations (Hellinga, 2019). While this suggests at least some generalizability, of course, some differences cannot be ruled out, for example masculinity, which is considered a major element in the profile of the typical fanatical soccer fan (Ben-Porat, 2010), could be more important and pronounced in some cultures than in some others. Lastly, the current study focused only on the subjective experiences and opinions of supporters and professionals. While exploring the topic from a behavioral point of view is important, this focus shed little light on for example health consequences and financial damage resulting from illegal use of fireworks inside the stadium. A study investigating the English Premier League shows that at least at English League Clubs, there is no longer an adverse connection between number of arrests and soccer club revenues (Jewell et al., 2014). Keeping in mind that hooliganism is often referred to as the *English Disease* (Green & Simmons, 2015), emphasizing the extent to which hooliganism and illegal fan behavior have been associated with English soccer, one might hence conclude that in the rest of the world financial consequences of illegal behavior inside the stadium are limited too. However, knowing that England took severe measures and invested heavily in enforcement of stadium rules, we cannot be sure if this lack of connection is applicable to clubs in

other countries. However, even if soccer clubs do not suffer large financial strain due to firework offences, deployment of police officers alone costs many man-hours and a lot of money (Nohlen & van Harreveld, 2017). When it comes to damage to health, studies in fireworks related incidents inside stadium are, to the best of our knowledge, yet to be performed. However, research shows that use of pyrotechnics inside the stadium resulted in large a amount of nanoparticles in the air, potentially dangerous to the health of supporters, and even more so to the health of athletes (Pirker et al., 2020).

### **Conclusion**

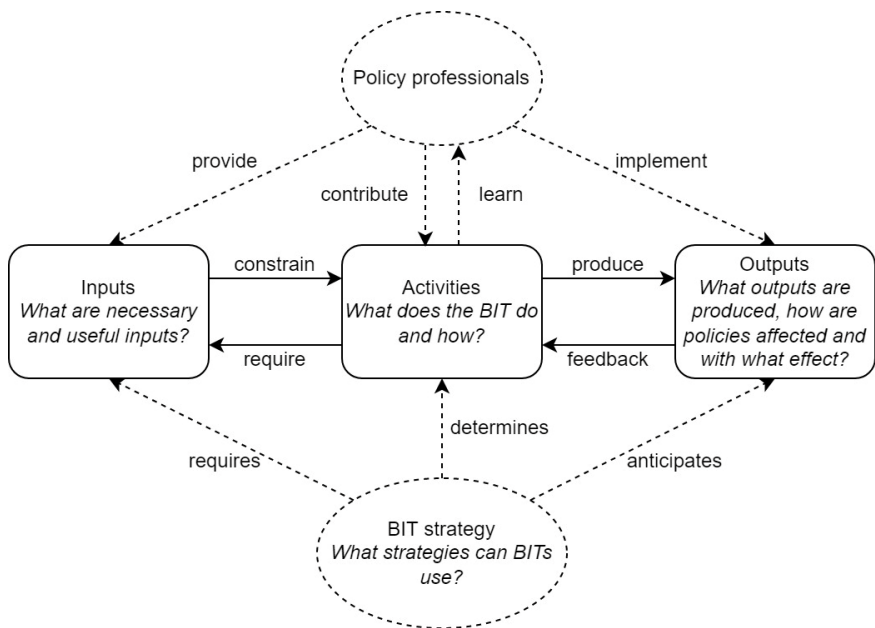
The current study aimed at gaining insight in the behavior of illegally lighting fireworks inside the soccer stadium by supporters. Qualitative interview data on the topic, provided by supporters and professionals were contextualized using insights from the hooligan literature, uncovering that although hooliganism and lighting fireworks inside the stadium share some underlying components, lighting fireworks is a distinct phenomenon, lacking the aggressive component that is central to hooliganism. When assessed from a behavioral perspective, motivation (reflective as well as automatic), as compared to capability and opportunity, seems to be the strongest facilitator of the behavior. This suggests that interventions directed at making mere changes in context are unlikely to have lasting effects on behavior. Supporters underline this notion by suggesting compromise and conversation as preferred future direction. Supporters feeling stigmatized by professionals could however be a barrier for this direction. However, since professionals contradict this view and talk mostly positively about fanatical supporters, building a bridge between supporters and professionals could be a promising starting point.

## Discussion

As outlined in the introduction, behavioural insights are an important resource for improving public policies by integrating the two. However, little is known about how such integration can be undertaken by specialised behavioural insights teams (BITs). Yet, BITs continue to be employed by numerous governments and other organisations around the globe. Therefore, the overarching purpose of the research presented in this dissertation was to investigate the integration undertaken by one BIT, the Behavioural Insights Group Rotterdam, (BIG'R). Together with my collaborators, I investigated the integration of behavioural insights into public policy from a conceptual perspective in Part 1. Crucially, BIG'R provided us with a unique opportunity to trial and research actual integration in practice over multiple years as reported in Part 2. Finally, we explored further potentials for integration in Part 3.

In this last chapter, I synthesise the insights from this research. First, the main findings are summarised and integrated into a single theoretical model. Subsequently, methodological considerations about the research and implications for practice are presented. Lastly, I outline suggestions for future research.

**Figure D.1**  
Theoretical model of BITs



**Theoretical Model of BITs**

Based on the research reported in this dissertation, I theorised a model describing the functioning of BITs (Figure D.1). To the best of my knowledge, this is the first model of its kind. The core of this model consists of activities of BITs to transform their inputs into outputs (depicted in solid lines). This part of the model is a simplification of the logic model for BIG'R reported in chapter 5. It is simplified to also be applicable to other BITs. However, the model is complemented with two external elements:



Chapter 6 and 8 showed that when interacting, BITs and external policy professionals mutually influence each other. Therefore, they are added as one element that is of importance but not directly part of BITs (e.g., as clients). In addition, BITs can vary and use differing strategies, as explained in chapter 8. Therefore, these strategies are added to the model as a second external element. In the following, I will describe the BITs model in more detail.

### ***Inputs***

The availability and quality of inputs enable but also constrain the activities of BITs. Behaviour is typically described, analysed, and changed using various perspectives and disciplinary traditions. Consequently, such perspectives and disciplinary backgrounds of the members of a BIT can influence a BIT's activities (Ewert et al., 2020). For instance, psychologists tend to put more focus on changing immediate choice contexts and the micro level for changing behaviour (e.g., reminders to eat healthy) when compared to sociologists who tend to address societal structures (e.g., dispersion of fast-food restaurants). In addition, there are diverse political perspectives on behaviour change. For instance, some political perspectives emphasize individual responsibility and freedom of choice whereas other perspectives suggest more responsibility for the state to steer and direct behaviour. Therefore, the range of disciplinary and political inputs are important for BITs. Note that this is an inclusive conception of inputs allowing diverse inputs rather than seeing the integration of behavioural insights into policy as determined by a single dominant ideology ("epistemic community"; Strassheim, 2021; Ewert, 2020; Feitsma & Whitehead, forthcoming).

As BIG'R embraced a collaborative working mode, it often required a substantial amount of effort and time to survey, negotiate, and integrate different perspectives from individuals with different backgrounds, as pointed out in chapter 1. The research results suggest though, that not different behavioural or political perspectives caused most disagreements within BIG'R, but rather divisions between scientists and practitioners pursuing different goals. Plausibly because within BIG'R the variability in disciplinary backgrounds was rather low (e.g., most researchers being psychologists) and most policy challenges concerned practical implementation challenges (e.g., increasing response rates for letters) rather than abstract political questions.

Further, more concrete limitations stemming from the available input were found in chapter 6, for instance the tenure of employees.

### ***Activities***

Main activities of BITs, including BIG'R, encompass efforts to address concrete policy issues, dissemination of behavioural insights, and capacity building for policy professionals (e.g., trainings). To improve and adapt their activities, BITs can look for feedback in the outputs that they achieve with their activities. For instance, BIG'R first exclusively focused on policy cases where it was possible to conduct a randomised control trial or another form of experimental evaluation, an approach similar to that of the BIT in Australia ("BETA"; Ball et al., 2021; Jones et al., 2021). However, this way BIG'R stayed away from some relevant policy challenges where a behavioural perspective may still have had

some added value. Ultimately, this motivated BIG'R to change its approach and focus on other policy cases as well. An example for such a project is reported in chapter 5, illustrating that more diverse policy challenges can be addressed by BITs once they refrain from conducting experiments only.

Another important feedback for BIG'R was that its policy proposals were often not implemented. To address this, BIG'R adapted its approach to include advocating and planning for implementation. This illustrates that a focus of BITs on "delivering" behavioural interventions (Hansen, 2018) is not enough. Rather, the model presented here assumes a shared responsibility between BITs and external policy professionals to translate behavioural insights into concrete and contextualised policy proposals and implement them, as explained in chapters 1 and 6. Actors of BITs are thus not just knowledge brokers (Feitsma, 2019a) but also action planners and implementers. As successful implementation is dependent on administrative processes and local policy priorities (Mintzberg & Water, 1985; Tummers & Bekkers, 2014), however, it is obvious that BITs require related inputs (e.g., policy expertise, political and administrative mandates) to achieve implementation and make a difference at scale.

### **Outputs**

BITs achieve a variety of outputs. As knowledge producers, they contribute novel behavioural insights; as intervention developers, they deliver effective interventions to change behaviour; as policy actors, they influence government actions and make a difference in individuals' lives. Often, these outputs overlap and are the result of a BIT addressing a single policy issue only. Consequently, the outputs from BITs can often be evaluated on multiple dimensions (e.g., knowledge dimension, welfare dimension; Kusters & van der Heijden, 2015).

The multitude of outputs is also reflected in the research design and findings of this dissertation. Chapters 2 to 5 contribute novel behavioural insights about behavioural determinants and/or test the effectiveness of behavioural interventions. From a policy perspective, however, chapters 2 to 5 may be considered pilots that still need to be translated and implemented into policy contexts. The general procedure and ingredients that BIG'R used for this translation and implementation are outlined in chapter 1, and in chapter 6 the effects of this procedure are investigated for all policy issues that BIG'R addressed.

In the following, I will illustrate the main knowledge gains from chapters 2 to 5 to illustrate some major outputs of BITs and BIG'R in particular. The improved repayment letter evaluated in chapter 2 shows that behavioural insights can be more than mere technical tweaks to increase the effectiveness of government actions. Rather, behavioural insights provide tools to better understand behaviours and reconsider tactical choices or policy priorities. In this case, the tactical choice was to stimulate contacting rather than enforcing repayment because this was more respectful towards the debtors, and debtors were assumed to be more responsive to that tactic. The nudges to increase compliance with an information security policy evaluated in chapter 3 were ineffective. Yet, the results point to some relevant conditions for nudges to work, namely that they should correspond to pre-existing attitudes. The third example about illegal garbage disposals reported in chapter 4 exemplifies that

nudges used to complement standard policy measures can boost the effectiveness of those measures with little effort and at low cost. The interviews about containing the use of fireworks in a football stadium reported in chapter 5 again show the potential of using behavioural insights for reconsidering tactical policy choices. Moreover, it serves as an illustration that the integration of behavioural insights into public policy is not limited to the field-testing of interventions, let alone nudges, and that behavioural insights can also be applied early during the policy cycle for problem exploration and policy formulation.

### ***Policy Professionals***

BITs do not function in isolation but are reliant on external policy professionals to provide their inputs, contribute to their activities (e.g., conducting experiments), and implement solutions developed in collaboration with the BIT. As a result, policy professionals and BITs interact regularly. Such interactions provide policy professionals with learning instances influencing their knowledge about behaviour and effective policy tools, as shown in Table 6.3. For instance, some professionals learned that behaviour can be programmed purposefully and does not need to be taken as a given. The knowledge of professionals about behaviour, in turn, influences their policy preferences, as illustrated by the findings from chapter 7 showing that policy professionals assuming behaviour to be fixed rather than malleable prefer more government intervention when aiming to change behaviour. Hence, taken together, policy professionals can learn from the interaction with BITs that behaviour can change and in turn assume fewer government intervention necessary to change behaviour.

This suggests that the integration of behavioural insights into public policy also takes place in the minds of policy professionals. What and how policy professionals think about behaviour and policy measures to change it does affect their decisions. Consequently, rather than only focussing on interventions to integrate behavioural insights into public policy, there is potential for addressing the cognitions of policy makers in the future.

### ***Strategies***

BITs implicitly or explicitly rely on strategies to align their functioning with a particular purpose and orchestrate their translation of inputs into outputs. The variety in strategies and corresponding purposes is shown in chapter 8. It highlights that BITs are not necessarily a uniform phenomenon dominated by common “frontstage models” and narratives (Ball & Feitsma, 2020; Feitsma, 2018), but rather a diverse group of different BITs which are all involved in the integration of behavioural insights into public policy in some sort (Strassheim, 2021). This also corresponds to the finding that recently practices of behavioural public policy started to become more diverse (Strassheim, 2020b).

BITs embracing different strategies implies some degree of self-awareness about what they are trying to achieve and how (e.g., generate behavioural insights by conducting research). This self-awareness shapes the link between the activities of BITs and related outputs on the one side and public policy and society as a whole on the other side.

## Methodological Considerations

Naturally, each individual part of this research has its methodological strengths and weaknesses. These are reported in the respective chapters and will not be repeated here. Instead, I will focus on general methodological considerations related to pragmatism, collaboration, open science, and ecological validity.

### *Pragmatism*

Different from many other research projects, this research was not driven by a fixed paradigm and a pre-determined research plan. Following Kuhn (1996), a research paradigm consists of shared beliefs (e.g., theories) from groups of researchers who agree about which questions are most meaningful and which methodology (e.g., experiments) is most appropriate for answering those questions. In contrast to this, the research reported in this dissertation was planned in an iterative process and in interaction with policy professionals from the municipality. This also meant that needs and constraints from policy practice could influence the choice of research methods and questions. For example, the needs from actors within the municipality could influence which policy issues were addressed by BIG'R to ensure their practical relevance, and administrative set-ups within the municipality could determine whether randomisation was possible thereby constraining methodological options (e.g., requiring quasi-experimental designs).

Acknowledging needs and constraints from practice encouraged an eclectic approach choosing and combining different research paradigms and methods ("multiple method research"; Johnson & Onwuegbuzie, 2004; Johnson et al., 2007). For instance, various data collection methods were used, ranging from standardised surveys to in-depth interviews. In addition, different analytic strategies and experimental designs were employed, including Bayesian inference, frequentist statistics, and content analysis. Moreover, some parts of the research were more exploratory whereas others were more confirmatory.

Reports from other BITs suggest that an eclectic approach separates BIG'R from many other BITs. Most BITs seem to analyse and change behaviour using pre-determined perspectives and tools. For instance, the handbook on behavioural insights from the OECD (2019) analyses behaviours using four major categories of behavioural determinants: attention; belief formation; choice; and determination. In addition, many BITs consider nudges their main intervention for changing behaviour (Ball, forthcoming; Ball & Feitsma, 2020) with BITs sometimes being called "nudge units" (Halpern, 2015). BIG'R was thus to some extent a rare BIT.

An eclectic approach corresponds to what has been labelled a "pragmatic" approach to research (Morgan, 2007; 2014). Following this approach, researchers are less concerned with abstract philosophical systems about knowledge and methodological commitments, but engage in contextualised and experience-based forms of inquiry and action that accept different methodological choices for various purposes. With a pragmatic approach, differences between paradigms are of lesser importance and ambitions to *prove* any paradigm become obsolete. Rather, the ambition is to engage in purposeful forms of inquiry. A pragmatic perspective implies that the popularity of

behavioural insights does not prove the superiority of any specific behavioural paradigm (e.g., the idea of steering behaviour to approximate rational behaviour) or innovations flowing from such paradigms (e.g., nudging), but that this popularity merely represents an opportunity of policy professionals becoming more familiar and open to *any* behavioural thinking and methods (Chetty, 2015).

The first main advantage of a pragmatic approach is that with more methodological flexibility, BITs can address policy challenges in a more contextualised way. As mentioned earlier, randomised control trials (RCTs) are central for the work of many BITs and they use them to evaluate their interventions (Ball, 2021; Einfeld, 2019; Haynes et al., 2012). While RCTs are a powerful tool for comparing the effect of different interventions in a specific context, they provide little information about the usefulness of interventions when being implemented elsewhere (Cartwright, 2010). Interventions that are effective in one context are not necessarily effective in other contexts as well.

Therefore, it has been argued that RCTs should better be embedded in a cumulative program that combines RCTs with other methods aiming to explain *why* interventions work and not just *if* they work (Deaton & Cartwright, 2018). An understanding of the causal mechanisms underlying interventions is central to realistically predict the effectiveness of those interventions in novel contexts and adapt them if needed. Many diverse methods can be used to investigate causal mechanisms (Marchionni & Reijula, 2019). For instance, the open-answer format used in the survey reported in chapter 3 helped to investigate causal mechanisms preventing the nudges from being effective. Once BITs are more open to diverse methods for learning about causal mechanisms, they can improve their ability to choose and adapt interventions. An obsessive fixation on RCTs on the other hand represents an unnecessary handicap.

A second main advantage of a pragmatic approach is that it allows BITs to address a larger scope of policy issues. Not every issue or intervention lends itself to being investigated using a RCT-approach. For instance, randomisation is often imperfect or impossible in real-life, as was the case for the interventions tested in chapters 2 and 3. However, analytic methods have been developed that allow rigorous research and inferences about causality without randomisation (e.g., Imbens & Wooldrige, 2009). Using such methods, BITs do not need to shy away when RCTs seem impossible. For instance, RCTs tend to be less useful than other methods during earlier stages of the policy cycle. To define and explore problems it is often useful to investigate opinions and preferences keeping an open mind (e.g., Árvai & Gregory, 2021). For this purpose, interviews are a valuable research method as exemplified by the interviews with football supporters in chapter 5. Information on preferences is difficult to obtain using RCTs only, however, and less empowering than interviews.

For behavioural public policy and strongly related disciplines such as psychology and behavioural economics, a pragmatic approach remains rather outlandish. This has some important practical consequences. First, researchers embracing a pragmatic approach need to become methodological generalists who are able to navigate different research paradigms and apply different methods. This requires a substantial amount of learning and research training. Second, multiple methods research is often more difficult to get published in scientific outlets than conventional and mono-methodological research. Third, pragmatic research as presented here in this thesis is also

more concerned with solving problems from practice rather than testing theory, which can further diminish the chances of publication in some scientific journals. Yet, this has been acknowledged and scientific traditions and institutions are changing, also at Erasmus University Rotterdam, to assign more value to applied research and societal impact (Smith et al., 2020; Smith et al., 2011). Nevertheless, for behavioural insights to reach their full potential at scale, there remain important methodological and research barriers to overcome.

### ***Co-Creating***

This research was undertaken as embedded research (Ward et al., 2021a; Ward et al., 2021b). Consequently, I was a full member of BIG'R and embedded in the municipality administration (e.g., having a desk at municipality office buildings). Such research encourages dynamic and fluid interactions (Ward et al., 2012) between researchers and policy professionals. It has been suggested that such collaborative efforts are better suited to close the research-practice gap than traditional and linear “push” or “pull” models (Marshall et al., 2014; Walter et al., 2003); with “push” models referring to efforts of knowledge producers for their knowledge to be implemented, and “pull” models referring to efforts of knowledge users to receive actionable evidence.

The research in this dissertation may be considered an example of engaged scholarship (Beaulieu et al., 2018; Boyer, 1996). According to Van de Ven and Johnson (2006), engaged researchers 1) confront real-life challenges, 2) organise the research as a collaborative endeavour of researchers and practitioners, 3) examine not only theoretical explanations but also alternative practical formulations, and 4) frame research findings to contribute to both knowledge production and domains of practice. Going beyond mere applications of academic knowledge (Greenhalgh & Wieringa, 2011), engaged scholarship within BITs thus brings together researchers and policy professionals to co-create novel ideas and solutions to advance both academic knowledge and public policy. Co-creation is considered a deeper, more meaningful, and socially respectful way of addressing policy issues (Cairney et al., 2016; Weible et al., 2012). It should be noted, however, that there are many barriers to successful co-creation and engaged scholarship, for instance power-imbalances and threats to scientific autonomy (e.g., Flinders et al., 2016), and difficult to navigate individual trade-offs for the involved researchers (e.g., Cairney & Oliver, 2020). In chapter 1, the co-creation process between researchers and practitioners embraced by BIG'R and some remedies for related challenges are discussed in detail.

Re-framing BITs as involved in co-creation can soften some concerns about technocracy and paternalism related to BITs and the use of nudges. As pointed out by John and Stoker (2019), BITs stand in a long tradition of science-led policy initiatives where scientists frame problems and offer related policy solutions, undermining individual autonomy of those targeted by those solutions. In fact, the strong emphasis on evidence-based practices from BITs and the seemingly indisputable nature of findings from RCTs facilitate enhanced authority and influence of researchers in the policy arena (Strassheim et al., 2015).

Taking a co-creating approach, however, implies an iterative exchange where researchers and policy professionals engage in a shared process of meaning-making that, at times, can be messy due to the complicated nature of policy processes and the versatility (or context-dependency) of behavioural findings (e.g., Feitsma, 2020; van Ryzin, 2021). From that perspective, the researcher perspective is added as just another input for public policy. It may even be argued that by aligning research with practical needs, co-creation helps to democratise academic knowledge that would otherwise remain isolated within academia. Yet, as John and Stoker (2019) point out correctly, involving those more during co-creation whose behaviour is targeted by BITs may improve adhering to norms of transparency and openness, for instance using the idea of self-nudging (Reijula & Hertwig, 2022). Chapter 5 exemplifies how BIG'R enabled the actual involvement of policy targets in the process of policy formulation.

### ***Open Science***

This research actively embraces practices of an open science culture. Open science aims to make scientific processes and results transparent and accessible (Nosek et al., 2015). It includes making research materials, hypotheses, statistical programming, and data freely available at appropriate time points. Such practices are said to ease replication leading to more reliable science and preventing further replication crises (Munafò et al., 2017). For instance, the hypothesis for multiple parts of this research had been pre-registered (Nosek et al., 2018) and the research material was made available online. If possible, the research was published using an open access format to be freely available online. Finally, the data and statistical programming was published online wherever possible.

### ***Ecological Validity***

Almost all the research of this dissertation was conducted in the field studying real-world phenomena, with chapter 7 being the only exception. As a result, this research provides a unique and in-depth account of the integration of behavioural insights into public policies as practiced within government. Being an insider to BIG'R provided me with fine-grained experiences and observations from practice that informed the design and conduction of this research (Atkins & Wallace, 2012; Brannick & Coghlan, 2007). All these aspects contribute to higher levels of ecological validity (Holleman et al., 2020). BIG'R being studied over an intensive period and from various perspectives probably makes it one of the most transparent and researched BITs so far.

The transparency and the extensiveness applied to the research enable readers to assume a critical attitude and consider similarities with their situation and whether those similarities warrant generalisation ("naturalistic generalisation", Stake & Trumbull, 1982). In fact, I encourage these considerations because even though the ecological validity is increased by the natural setting of this research, ecological validity is low when findings are difficult to generalise because of some unique features of BIG'R potentially limiting generalisation.

## Implications and Suggestions for Practice

The municipality of Rotterdam played an important role by providing the means and context to conduct the research reported here. Consequently, next to some global implications of this research, I outline some suggestions for the municipality to further improve the integration of behavioural insights into Rotterdam public policy.

### *Global Implications*

This dissertation provides positive evidence for the employment of BITs if three conditions are met. In the following, I will describe those conditions that I believe are useful for policy professionals.

There is overwhelming evidence that using a realistic understanding of behaviour for formulating and implementing public policies improves those policies, and that behavioural interventions developed by BITs are effective (e.g., DellaVigna & Linos, 2020; Thaler & Sunstein, 2008). Chapters 2, 4, and 5 corroborate this evidence in describing effective interventions and/or useful policy advice generated by BIG'R. Therefore, organisations may continue to employ BITs when aiming for behavioural input that contributes to solving local policy challenges. However, this comes at the cost of providing the necessary resources for the BIT and yet having no guarantee that the BIT will develop effective solutions because, as this research also shows, forming a BIT is not a straightforward task and solutions developed by BITs can be ineffective (see also Hummel & Maedche, 2019; Van Ryzin, 2021). Consequently, BITs are plausibly the most useful when tasked with policy challenges that are relevant enough to justify spending extensive resources for addressing them, but at the same time not so relevant that failures to effectively address the challenges are intolerable. However, even when failures are intolerable, also attending to behavioural perspectives can help to prevent some failures related to naive misconceptions about behaviour. Something that has been exemplified during the Covid-19 pandemic multiple times (e.g., Feitsma & Whitehead, 2021; Harvey, 2020; Mahase, 2020; Michie et al., 2020).

BITs are not the only format for integrating behavioural insights into public policy. As pointed out in the introduction, baseline capacity building for large groups of an organisation's members may also be an appropriate format and, in fact, behavioural insights continue to be popular among policy professionals during the "age of behavioural sciences" (Sunstein, 2016) even without BITs. An example for this may be seen in the approach of the Obama administration in the United States evolving around an Executive Order requiring behavioural insights to be considered and applied as a standard for all public policy (Congdon & Shankar, 2015; Executive Order No. 13707, 2015). The format of capacity building can be applied at scale much easier than employing BITs: Every public servant can be trained about behavioural insights, but it is impossible to have a BIT ready to advise on every policy issue. This suggests that BITs are useful when tasked with issues of high behavioural complexity that cannot be made sense of using standard behavioural understanding. This is also reflected in chapter 1 where involving BITs is framed as an additional perspective that enhances the problem-solving capacity of organisations.



In the past, BITs played an important role in accelerating the popularity of behavioural insights and innovating their integration into public policy. After more than three years, more Rotterdam public servants knew BIG'R than they knew behavioural insights for instance. In addition, the BIT UK has had a huge influence on the way that behavioural insights are currently practiced (e.g., Ball & Feitsma, 2020; Jones et al., 2021). BITs have thus an important signalling and educative function influencing the way that behavioural insights are practiced within one's organisation and beyond. This suggests that BITs are useful when employed strategically to innovate and determine the integration of behavioural insights into public policy within one's organisation.

To summarise the above, BITs are useful when 1) employed for challenges of medium relevance, 2) underlying behaviours cannot be made sense of using standard approaches, and 3) they are part of a strategy to improve the integration of behavioural insights into public policy.

To design and develop BITs purposefully, organisations can make use of the strategic dimensions presented in chapter 8 as reflective prompts. Moreover, policy professionals may benefit from the propositions resulting from the evaluation of BIG'R presented in Table 6.5. Importantly, I recommend policy professionals to pay attention to the translation of behavioural insights into successful policy programmes. Successful programmes are concerned with translating evidence about effective interventions into action plans that can be implemented at scale and with positive effects. Often, BITs seem to be focused on the aspect of effectiveness neglecting the implementation challenge and that translational processes and actions require time, money, and effort (Damschroder et al., 2009).

The last global implication targets educational institutions, namely universities. Multiple aspects mentioned suggest revisiting the way some behavioural professionals are educated. For behavioural professionals to succeed at BITs, education needs to equip them with the mindset and skills to engage in participatory working modes, to think creatively and pragmatically about research methods, to collaborate with individuals from other disciplines, to understand the policy context, and to translate behavioural evidence into policy programmes. Yet, related knowledge and skills are rarely taught within related programmes (e.g., psychology, economics), requiring educational institutions to revisit their programmes.

### ***Local Suggestions***

The municipality, at the end of the pre-determined project period for BIG'R, decided to not continue BIG'R in its original form but instead facilitate an internal network of policy professionals interested or involved in behavioural insights to further advance the integration of such insights into Rotterdam public policy. Such a network, intended to accommodate former members of BIG'R, may be described as a community of practice (Li et al., 2009; Wenger, 1998), advocacy coalition (Sabatier & Weible, 2007), epistemic community and/or an instrument constituency (Strassheim, 2021) depending on the concrete design and purpose of the network. Communities of practice are mostly focussed on learning, advocacy coalitions on policy change, epistemic communities on shared policy-relevant knowledge, and instrument constituencies on promoting policy instruments. At the time of writing, the

design and purpose of the network were still in development. Given the different foci briefly sketched here, the first suggestion is to specify the network's mission, design, and objectives, and evaluate the required resources. Crucially, research suggests that all these networks require relevant administrative, political, and financial resources to reach their objectives.

The findings from this research suggest that implementation of interventions and policy proposals developed by BIG'R can be improved. Although BIG'R regularly advocated implementation, its resources and capacities to stimulate implementation were limited as outlined in chapter 6. BIG'R did, for instance, not have the political support that the BIT UK had in its beginning by being based in the UK Cabinet Office (John, 2014). Acknowledging that implementation is an effortful process rather than a one-time event (Damschroder et al., 2009; Nilsen, 2015), I suggest a programme to synthesise the most promising insights and interventions from BIG'R given current policy priorities and to plan and conduct their implementation.

The third suggestion concerns aspects of knowledge mobilisation. The policy professionals involved in BIG'R acquired technical as well as conceptual knowledge (Katz, 1955) about the integration of behavioural insights into public policy. This knowledge concerns for instance knowledge about behaviour change techniques, or procedures for integrating behavioural insights into public policy. In addition, many of the policy professionals have experience in giving trainings to their colleagues about behavioural insights. The municipality can leverage on this experience by institutionalising knowledge mobilisation (Ward, 2017). Specifically, they may adapt the job descriptions of the policy professionals involved in BIG'R to (also) function as behavioural experts and consultants within the administration and/or they may institutionalise capacity building activities (e.g., trainings) delivered by those individuals. Such activities can be informed by the needs that became apparent during the project period of BIG'R, specifically introductory and inspirational sessions on the one hand, and more extensive trainings enabling policy attendees to apply behavioural insights on their own on the other hand.

## **Directions for Future Research**

I will conclude this dissertation with three recommendations for future research. The first suggestion is to further investigate the integration of behavioural insights into public policy, the second is to pay more attention to the contexts that BITs are operating in, and the last suggestion is to continue the development of effective behavioural interventions for public policy.

### ***Integration***

The integration of behavioural insights into public policy was re-fuelled with the popularity of nudges (Thaler & Sunstein, 2008). Nudges even became somewhat synonym for that integration and BITs were labelled "nudge units" (Halpern, 2015). After the introduction of nudging, there were high expectations (and fears) associated with behavioural insights from both scientists and practitioners alike (e.g., Furedi, 2011; Marteau et al., 2011). Such a euphoric initiation phase is not new for

innovations, as are subsequent phases of disappointment and readjusting expectations (Dedehayir & Steinert, 2016).

Although some readjustments of expectations have taken place with regard to nudging (e.g., Selinger & Whyte, 2012; Thaler & Sunstein, 2021), the integration of behavioural insights into public policy is in general still in an early phase and perceived rather novel. Even after four years of BIG'R, most municipality employees were unaware of behavioural insights being integrated into public policy. Therefore, the first recommendation for future research is to continue both innovating and investigating behavioural insights and BITs. By venturing novel formats of BITs that differ from current examples, the full potential of behavioural insights can be explored and leveraged. In terms of research, historical analyses may be used to explore how BITs evolve, adapt, and may shape the future (for an inspiring case see Mukherjee & Giest, 2020), and comparative studies (e.g., Jones et al., 2021) across BITs may be used to distil best practices and lessons learned for other BITs to benefit from.

### **Contexts**

The popularity of nudging and behavioural insights means that governmental doors were opened, both literally and figuratively, for behavioural professionals to walk through and penetrate the policy arena more than ever before. Above, I have outlined some consequences of this for the behavioural sciences (e.g., a need for other skills and methodological approaches).

With behavioural professionals and policy professionals interacting, however, behavioural professionals also affect the policy arena. This becomes evident in updated and novel tools that policy professionals may use (e.g., nudging), changed practices of policy development (e.g., experimentation), and altered cognitions of policy professionals about behaviour and policy in general (e.g., seeing behaviour as fixed or malleable). All of which may ultimately change the role of government and state-citizen relations (e.g., Leggett, 2014; Strassheim, 2020a). Although there are some notable exceptions (e.g., Hampton & Adams, 2018; Whitehead et al., 2017), empirical investigations into the effects that behavioural insights have on the policy arena are rare. Therefore, I recommend investigating those effects more, plausibly with policy professionals that have collaborated with BITs as major informants since most past research has focussed on behavioural professionals themselves.

With more behavioural scientists entering government, they pay increasing attention to the behaviour and decisions of policy professionals. Evidence indicates that those policy professionals are subject to the same biases and heuristics as other individuals (e.g., Lodge & Wegrich, 2016; Vis, 2019). They are of bounded rationality. In fact, the BIT UK recently published a report about biased decision-making of policy professionals (Hallsworth et al., 2018). This perspective on policy making is also adopted in recent studies that investigate ideal ways of presenting information to policy professionals (Baekgaard et al., 2019), or re-framing strategies for policy professionals (Nagtegaal et al., 2020). Both examples see the decision-making of policy professionals as fallible and use psychological knowledge to simulate better decisions. Such research answers the more general

question of how public institutions can be designed from a psychological perspective to ensure good governmental outcomes. This question may be extrapolated: How does a state look like that was designed by behavioural scientists (and not the ancient Greeks, economists, lawyers, administrators, etc.) for optimal governmental outcomes? Behavioural scientists have only begun to develop ideas for a psychological model of the state (e.g., Thaler & Sunstein, 2008) and I recommend that they continue.

### ***Effective Interventions***

BIG'R has developed and exemplified some effective interventions to be used by governments. Evaluations of such interventions are reported in chapters 2 and 4. For scientists, the development and testing of such interventions in the field provides a rare opportunity to achieve positive societal impact. Such researcher-practitioner partnerships to develop interventions, either facilitated by a BIT or not, continue to be of high relevance and importance for both researchers and practice (Smith et al., 2020; Smith et al., 2011). Therefore, my last recommendation for future research is to continue the development and trialling of interventions using behavioural insights. In addition, however, I encourage such researchers to adopt a mindset that extends beyond the development of interventions to also include their implementation. For this, researchers may benefit from concentrated implementation efforts and insights from implementation science, as explained in chapter 6.

## Appendices

# Summary

For policy making, it is typically assumed that humans act rationally. Yet, a large body of behavioural research shows that many decisions and behaviours are not the product of rational and logical thought but influenced by, amongst others, heuristics, habits, emotions, and intuitions (Kahneman, 2011; Thaler & Sunstein, 2008). Therefore, a difference can be observed between rational actors and humans. The assumption for policy making of humans acting rationally is thus often wrong. For effective public policy, it is important to avoid such wrongful assumptions and adopt a more realistic understanding of behaviour (Thaler & Sunstein, 2008). To achieve this, numerous governments and other organisations during the last decade started exploring ways of integrating a more realistic understanding of behaviour (i.e., behavioural insights) into public policy.

A well-known example of policy making informed by behavioural insights are *nudges* (Thaler & Sunstein, 2008). Nudges are small and seemingly irrelevant changes in choice contexts that still influence behaviour. An example of a nudge is the new law for organ donor registration in the Netherlands that as a default assumes everyone to be an organ donor unless indicated otherwise. A rational choice agent would not be influenced in his or her choice whether to be an organ donor by the defaulted option and, yet, there are more organ donors when active opt-out rather than opt-in is required (Johnson & Goldstein, 2003). This is because individuals often assume the defaulted option to be the desired option and tend to prefer the easy option of inaction (Jachimowicz, 2019). Generally, most nudges are effective (Benartzi et al., 2017; DellaVigna & Linos, 2020; Hummel & Maedche, 2019) and they are often employed by governments and many other organisations.

To develop nudges and integrate behavioural insights into public policy, governments often employ specialised teams, so-called behavioural insights teams (BITs). BITs typically consist of policy professionals and behavioural experts to support governments and organisations with making realistic assumptions about behaviour for policy making. Most BITs follow in their approach the first BIT that was set up by the government of the United Kingdom in 2010 (Halpern, 2015). As a result, many BITs embrace an analytic approach, using scientific knowledge to investigate behaviours underlying specific policy challenges. In addition, those BITs develop interventions to change behaviour and test their interventions in the field using experimental methods (e.g., John, 2014; Einfeld, 2019). BITs are therefore related to the general trend of evidence-based rather than ideology-based public policy (e.g., Head, 2010).

Little is known, however, about how BITs best function and what their success factors are (Ball & Head, 2020; Gofen et al., 2021). For governments and organisations, this complicates starting up and developing BITs. Therefore, the research of this dissertation aimed to investigate the functioning of BITs and the research question was as follows: *How can behavioural insights be integrated centrally into public policy?*

To answer this question, I became a member of one BIT, the Behavioural Insights Group Rotterdam, for three years and I completed several related research projects. BIG'R existed for four years and it was the BIT of the municipality of Rotterdam and the Erasmus University Rotterdam. The research projects belong to three different parts that together form the core of this dissertation. Part 1 aims to describe the functioning of BIG'R from a conceptual perspective. Part 2 is about actual

integration of behavioural insights into public policy as practiced within BIG'R. Part 3 examines integration strategies to be applied in the future. The results of the different research projects enabled me to propose a model for the functioning of BITs as part of the discussion.

## Part 1

In **Chapter 1**, the approach of BIG'R to address policy challenges is described on a conceptual level. This approach relies on three kinds of inputs: policy expertise, behavioural research, and collaboration efforts. In addition, the approach entails a four-stepped procedure for combining these inputs in a meaningful way. The first step of this procedure is to choose relevant policy issues. In a second step, exploratory research is conducted to investigate related behaviours and the context of the policy issue. Based on the findings, behavioural interventions are then developed and trialled in the field whenever possible. For the third step, the ideas and if applicable the trial results are translated into policy advices that are shared with the municipality. Finally, for the fourth step, BIG'R supports the implementation of the advice. This approach was developed iteratively by BIG'R based on the available literature, (e.g., Ames & Hiscox, 2016; Service, 2014; Congdon & Shankar, 2015; Halpern, 2015; Halpern & Sanders, 2016; Haynes et al., 2012; Thaler & Sunstein, 2008), experiences from practice, and related theorising.

This chapter is a relevant contribution to the literature about BITs because it provides the first detailed description of the inputs required for BITs. Moreover, in this chapter it is explained how BIG'R addressed four common challenges for BITs (an overreliance on randomised control trials, a limited understanding of contexts, threats to good scientific practice, and bounded rationality of individuals applying behavioural insights). The chapter therefore also provides valuable insights and tips for other BITs.

## Part 2

Part 2 contains five chapters describing actual integration of behavioural insights into public policy within BIG'R: Chapters 2 to 5 concern specific policy issues that BIG'R addressed and the solutions it developed; chapter 6 reports an evaluation of BIG'R.

In **Chapter 2**, nudges serving to improve a repayment letter are described and evaluated in the field. Some Rotterdam residents received unjustified welfare payments (e.g., because of additional incomes that they reported too late for the municipality to adjust the payments) that those residents then needed to repay. For this, residents received a letter informing them about the repayment. Since too many of those residents did not react to the letter, BIG'R developed nudges to improve the letter. These nudges aimed to stimulate debtors to contact the municipality so that both parties could agree on repayment by instalments. This way, the nudges acknowledged that most residents receiving the letter plausibly experienced financial scarcity preventing them from being able to repay the debt immediately and in full. An evaluation, comparing the standard letter without nudges and the letter with nudges, showed that the nudges were effective. The likelihood of residents agreeing on repayment by instalments was 1.27 times higher with the nudging letter than with the standard letter. The chapter

thus shows that the approach of BIG'R can be effective and that to receive repayments it can be beneficial to focus on stimulating contacting behaviour and agreeing on repayment by instalments rather than repayment alone.

In **Chapter 3**, nudges were used to improve the compliance of municipality employees with a security policy. This security policy required employees of the municipality to wear an identifying lanyard around their neck in specific office areas so that they would be recognisable as authorised to be there, compared to others without a lanyard that could then be approached and asked about their authorisation. However, about half of the employees did not wear the lanyard. BIG'R therefore developed two nudges and tested them in the field. The nudges served to remind employees to wear the lanyard and to increase their motivation to wear it. Yet, the nudges proved ineffective and led to negative reactions from some employees. As a follow-up, BIG'R therefore conducted a survey among the employees showing that some employees disliked the nudges because they had little comprehension for the lanyard policy and did not agree with it. This chapter therefore exemplifies that for nudges to be effective, the target context and potential negative reactions need to be taken into account.

**Chapter 4** concerns nudges that were combined with a standard policy measure, this way better taking into account the context of the nudges. Those nudges served to improve the effectiveness of door-to-door canvassing aiming to reduce illegal garbage disposals. In some neighbourhoods of Rotterdam, residents did not dispose their garbage in dedicated street containers but next to them, for instance because container openings were too small for the garbage. The nudges were stickers that the employees conducting the canvassing handed over to residents at the end of their conversation. By placing the stickers on or close to their front doors, residents committed to keeping their street clean and free of garbage. At the same time, the stickers communicated a positive social norm for the whole neighbourhood that was also strengthened by signs put next to all street containers. The nudges proved to be highly effective, reducing illegal garbage disposals by two thirds. This chapter shows the potential of nudges when combined with and integrated into existing policy measures (see also de Ridder et al, forthcoming).

**Chapter 5** provides an example of integrating behavioural insights into public policy during the formulation phase of public policy making. Hence, it is not about any interventions, but general policy directions related to the use of fireworks by football supporters during matches in a Rotterdam football stadium ("De Kuip"). Using interviews with supporters, their drivers for lighting fireworks were investigated, as well as the beliefs of professionals which provided a frame of reference. The results showed that enhancement of the atmosphere inside the stadium and supporting the players on the field were amongst the most important drivers for supporters to light fireworks. From a behavioural perspective, it seemed that a strategy aiming to facilitate a dialogue and compromises between supporters and professions to regulate the use of fireworks was the most promising policy direction. This chapter shows the potential of behavioural insights to develop policy options, and it illustrates potential contributions of BITs that do not limit themselves to only conduct experiments.



In **Chapter 6**, BIG'R is evaluated comprehensively. For this, its inputs (e.g., members, budget), activities (e.g., projects, trainings), outputs (e.g., implemented policy advices), and policy outcomes (e.g., less garbage in the streets) were systematically described and investigated (Saunders et al., 2005). In addition, the context that BIG'R operated in was taken into account. Using surveys and interviews, multiple groups participated in this research, for instance members of BIG'R, proposers of policy cases, and training attendees. The results led to the formulation of nine propositions to describe the integration of behavioural insights into administrative operations (Table 6.5).

Chapter 6 is, to the best of my knowledge, the first scientific evaluation of a complete BIT. This way, it also investigates the translation of inputs into outputs and policy outcomes. Past evaluations mainly focussed on outputs of BITs, namely interventions (e.g., DellaVigna & Linos, 2020), but not complete BITs and how BITs may be embedded into organisations.

## Part 3

Part 3 consists of two chapters that point out possible future ways of integrating behavioural insights into public policy.

In **Chapter 7**, using a survey design, it is investigated how implicit assumptions of policy professionals about behaviour change influence policy preferences of those professionals. The results suggest that policy professionals assuming behaviour to be more fixed and difficult to change are more likely to prefer strict policy tools and information provision to change behaviour than policy professionals assuming behaviour to be malleable. A possible explanation for this is that for policy professionals more government intervention to change behaviour is required when individuals are assumed unable to change their behaviour themselves.

This chapter is one of the first studies investigating the influence of behavioural factors on policy preferences of policy professionals (Capano & Howlett, 2020; Howlett et al., 2020). It demonstrates that public policy is influenced by what policy professionals think about behaviour. Integrating behavioural insights into public policy is thus not only about interventions (e.g., nudges) and public policy itself but also about the minds of policy professionals. In the future, this insight may be exploited further, for instance by ensuring that policy professionals make realistic assumptions about behaviour when making policy choices.

**Chapter 8** is about different strategies that BITs can employ. Based on a systematic investigation of the experiences with BIG'R, three strategic dimensions for BITs are provided and explained. One dimension is about whether BITs use a scientific approach to contribute novel behavioural insights or a practice-oriented working mode to solve real-life policy challenges. Another dimension describes the difference between BITs that either aim to efficiently develop useful interventions or to teach and learn about how behavioural insights and public policy can be integrated. The last dimension concerns coordination among members of a BIT who either adhere to a strong hierarchy or a networked approach that functions without a strong hierarchy.

This chapter contains the first investigation of different strategies for BITs, also showing some of the variety in integrating behavioural insights into public policy. Moreover, practitioners can use the

dimensions as reflective prompts to make informed decisions about the type of research and outcomes that they expect from their BITs.

## Discussion

In the final chapter, the research question (How can behavioural insights be integrated centrally into public policy?) is being answered by combining the results from the previous parts and chapters into a theoretical model about the functioning of BITs (Figure D.1). The core of this model consists of inputs required by BITs, their activities, and the outcomes and effects that they achieve.

Inputs enable the activities of BITs. For BIG'R, the inputs were described in chapter 1 and evaluated in chapter 6. Similar to BIG'R, the main activity of many BITs is to investigate specific policy issues from a behavioural perspective and develop solutions to change behaviour. In addition, many BITs, including BIG'R, also engage in the training of policy professionals so that they can make use of behavioural insights themselves (e.g., Baggio et al., 2021). The outcomes of BITs (e.g., policy proposals) are ideally implemented by external policy professionals with positive effects. In practice, however, not all BIG'R proposals were implemented and some interventions from BIG'R proved to be ineffective (e.g., chapter 3). Therefore, BIG'R over time spend more and more effort on implementation.

The model acknowledges that external policy professionals play an important role for BITs to provide the necessary inputs (e.g., budget), to collaborate for the activities (e.g., sharing information, describing requirements for interventions), and to implement proposals and tools. In addition, policy professionals interacting with BITs also gain new experiences and learn about the importance of behavioural insights for more effective public policy. The model moreover acknowledges that BITs can employ different strategies influencing their required inputs, activities, and outputs. Therefore, external policy professionals and strategies of BITs are both included in the model. To the best of my knowledge, this is the first model about the functioning of BITs described in the scientific literature.

Subsequently, I discuss methodological aspects of the research. The approach of BIG'R and the research of this dissertation were based on a pragmatic approach to research (Morgan, 2007; 2014). Such a perspective is less concerned with abstract philosophical discussions about the nature of knowledge and which methodological tools are appropriate for producing that knowledge; it rather aims to engage in contextualised and experience-based forms of inquiry that can inform action. It allows researchers to combine different research methods (e.g., qualitative and quantitative) and to choose for each research question the most appropriate method or combination of methods. The main advantage of a pragmatic approach is that it enables BITs to address a larger variety of policy challenges than compared to a situation where BITs commit to only one type of research. So far, BITs often limited themselves to conducting randomised control trials (e.g., Jones et al., 2021). In contrast to this, chapter 5 illustrates that by conducting interviews BITs can make more diverse contributions to public policy.

Another essential element of the approach of BIG'R and multiple research projects of this dissertation, namely the research reported in chapters 2 to 5 about examples from actual integration,

was co-creation. The research partly originated and was conducted in collaboration with policy professionals, for instance to evaluate the relevance of policy issues that were proposed to BIG'R. Co-creation is often claimed to be an appropriate tool for bridging the gap between practice and scientific knowledge (Marshall et al., 2014; Walter et al., 2003). In addition, using co-creation helped to come up with policy proposals that take into account local contexts and non-scientific knowledge from practice.

Lastly, implications for practice based on the findings of this research and suggestions to further improve the integration of behavioural insights into Rotterdam public policy are discussed. The first implication for practice is that BITs are best employed to address policy issues that are sufficiently important to legitimise the resources invested in BITs but not so important that failing to address the issue is not an option. This is because the research shows that solutions developed by BITs can be effective but are not always so. The second implication is that BITs best focus on policy issues that cannot be addressed effectively using standard policy approaches. Behavioural insights are currently spreading, and policy professionals become more competent in making use of those insights. The competencies of BITs are an addition to that, best used for the most complex behavioural issues. Third, BITs should be embedded in a strategy to improve the integration of behavioural insights into public policy, for instance to be able to assign clear objectives to BITs. The suggestions for the municipality of Rotterdam concern the idea of using a formal network of policy professionals to further deepen and spread knowledge about behavioural insights within the municipality, the implementation of findings from BIG'R, and the mobilisation of the knowledge acquired by former BIG'R members.

Finally, three promising directions for future research are discussed. The first direction concerns researching BITs further to better understand what contributes to them being successful. The second direction is to investigate the consequences of BITs on their policy context. The last direction is to continue the development of effective behavioural interventions for public policy.



## Nederlandse samenvatting (Summary in Dutch)

Bij het maken van beleid wordt er doorgaans van uitgegaan dat burgers rationele beslissers zijn. Veelvuldig onderzoek laat echter zien dat gedrag en beslissingen vaak niet het resultaat zijn van rationele overwegingen en logica, maar eerder van onder andere vuistregels, gewoontes, emoties en intuïties (Kahneman, 2011; Thaler & Sunstein, 2008). Hierdoor is er vaak een verschil tussen een rationele beslisser enerzijds en menselijk gedrag anderzijds. De aanname dat burgers rationele beslissers zijn klopt dus vaak niet. Voor regeringen en andere organisaties is het belangrijk om foutieve opvattingen over gedrag te vermijden en om uit te gaan van een realistisch beeld van gedrag om effectief beleid te maken (Thaler & Sunstein, 2008). Tijdens het laatste decennium zijn daarom talrijke regeringen en andere organisaties begonnen met het zoeken naar manieren om een realistisch mensbeeld (dat wil zeggen gedragsinzichten) te kunnen combineren met openbaar beleid.

Een bekend voorbeeld van het maken van beleid met gedragsinzichten zijn *nudges* (Thaler & Sunstein, 2008). Nudges zijn kleine veranderingen in keuzeomgevingen die irrelevant lijken maar het gedrag van mensen toch beïnvloeden. Een voorbeeld hiervan is de nieuwe donorwet, waarbij iedereen standaard als orgaandonor wordt geregistreerd tenzij iemand actief een andere keuze doorgeeft. Een rationele beslisser zal in zijn/haar keuze om wel of niet orgaandonor te zijn niet door de standaardinstelling worden beïnvloed. Toch zijn er door de nieuwe donorwet meer mensen orgaandonor (Johnson & Goldstein, 2003). Dit heeft ermee te maken dat mensen de standaardinstelling vaak automatisch als de gewenste optie zien en soms te gemakzuchtig zijn om actie te ondernemen (Jachimowicz, 2019). In het algemeen zijn nudges meestal effectief (Benartzi et al., 2017; DellaVigna & Linos, 2020; Hummel & Maedche, 2019) en worden ze inmiddels regelmatig toegepast door overheden en veel andere organisaties.

Om nudges en beleid met gedragsinzichten te ontwikkelen, maken veel overheden gebruik van hierin gespecialiseerde teams, zogenoemde gedragsteams ("behavioural insights teams"). Gedragsteams bestaan normaliter uit beleidsmedewerkers en gedragsdeskundigen en helpen regeringen en organisaties om bij het maken van beleid uit te gaan van een realistisch beeld van menselijk gedrag. De meeste gedragsteams volgen met hun werkwijze het voorbeeld van het eerste gedragsteam. Deze werd opgericht in 2010 door de overheid in Groot-Brittannië (Halpern, 2015). Veel van deze gedragsteams volgen daarom een aanpak die gebaseerd is op een analytische benadering en wetenschappelijke kennis om concrete gedragingen die ten grondslag liggen aan een specifiek beleidsprobleem te onderzoeken. In een tweede stap worden dan interventies die het gedrag moeten veranderen ontworpen en getest door middel van veldexperimenten (e.g., Einfeld, 2019; John, 2014). Gedragsteams zijn daardoor aangesloten bij de algemene trend om kennisgedreven beleid te maken in plaats van ideologiegedreven beleid (bijvoorbeeld Head, 2010).

Er is echter nog weinig bekend over hoe gedragsteams het beste werken en wat ervoor kan zorgen dat die teams succesvol zijn (Ball & Head, 2020; Gofen et al., 2021). Dat maakt het moeilijk voor regeringen en organisaties om gedragsteams op te richten en te ontwikkelen. Het doel van dit proefschrift was daarom om de werkwijze van gedragsteams te onderzoeken en de onderzoeksvraag luidde: *hoe kan je door middel van gedragsteams beleid maken met gedragsinzichten?*

Om de vraag te beantwoorden was ik voor drie jaar onderdeel van een gedragsteam, de Behavioural Insights Group Rotterdam (BIG'R), en heb ik verschillende onderzoeken uitgevoerd. BIG'R bestond vier jaar en was het gedragsteam van de Gemeente Rotterdam en de Erasmus Universiteit Rotterdam. De verschillende onderzoeken zijn samengevoegd in drie secties die in het volgende beschreven worden. Sectie 1 heeft als doel de werkwijze van BIG'R te beschrijven op een conceptueel niveau. Sectie 2 omvat voorbeelden uit de praktijk van BIG'R die laten zien hoe je beleid kan maken met gedragsinzichten. In sectie 3 wordt op basis van de ervaringen met BIG'R onderzocht hoe je toekomstig beleid zou kunnen maken met gedragsinzichten. De resultaten van de verschillende onderzoeken stellen mij in staat om in de discussie een model op te stellen dat de werkwijze van gedragsteams beschrijft.

## Sectie 1

In **Hoofdstuk 1** wordt de werkwijze van BIG'R om beleidsvraagstukken op te pakken beschreven op een conceptueel niveau. Deze werkwijze maakt gebruik van drie ingrediënten: beleidsexpertise, gedragsonderzoek en samenwerkingsinspanningen. Verder omvat de werkwijze een procedure om die ingrediënten op een zinvolle manier te combineren. De procedure omvat vier stappen en start met het oppakken van een relevant beleidsvraagstuk. In de tweede stap wordt exploratief onderzoek verricht om bijbehorende gedragingen te onderzoeken en de context van het probleem te schetsen. Op basis daarvan worden ideeën voor gedragsinterventies bedacht en waar mogelijk getest in het veld. In een derde stap worden de ideeën en, indien van toepassing, de testresultaten vertaald naar een beleidsadvies dat wordt overhandigd aan de gemeente. In een vierde stap steunt het gedragsteam de implementatie van het advies. Deze werkwijze werd door BIG'R ontwikkeld in een iteratief proces op basis van de bestaande literatuur (e.g., Ames & Hiscox, 2016; BIT, 2014; Congdon & Shankar, 2015; Halpern, 2015; Halpern & Sanders, 2016; Haynes et al., 2012; Thaler & Sunstein, 2008) en eigen ervaringen en reflectie.

Dit hoofdstuk is een belangrijke bijdrage aan de literatuur over gedragsteams omdat het voor het eerst de ingrediënten voor het maken van beleid met gedragsinzichten uitgebreid beschrijft. Daarnaast wordt in het hoofdstuk uitgelegd hoe BIG'R om is gegaan met vier vaak voorkomende uitdagingen voor gedragsteams (te grote focus op gerandomiseerde onderzoeken met controlegroep, onvoldoende aandacht voor contextuele factoren die gedrag beïnvloeden, bedreigingen voor wetenschappelijke nauwkeurigheid, beperkte rationaliteit van mensen die beleid maken om rationeel gedrag te stimuleren). Op deze manier voorziet het hoofdstuk ook andere gedragsteams van waardevolle inzichten en tips.

## Sectie 2

Sectie 2 omvat vijf hoofdstukken die de praktijk van BIG'R beschrijven: Hoofdstukken 2 tot en met 5 gaan daarbij over specifieke beleidsvraagstukken die BIG'R heeft opgepakt en de oplossingen die door BIG'R zijn bedacht en onderzocht; hoofdstuk 6 omvat een evaluatie van BIG'R.

In **Hoofdstuk 2** worden nudges beschreven en geëvalueerd die zijn ontwikkeld om een terugvorderingsbrief te verbeteren. Soms kregen bewoners te veel uitkering uitbetaald (bijvoorbeeld omdat ze extra inkomsten te laat hebben gemeld aan de gemeente om de uitbetalingen aan te kunnen passen) die zij dan moesten terugbetalen. Daarover kregen de bewoners een terugvorderingsbrief. Omdat veel ontvangers niet op de brief reageerden heeft BIG'R nudges ontwikkeld en deze vervolgens toegepast in de brief. Dit had als doel om meer bewoners na ontvangst van de brief contact op te laten nemen met de gemeente om een betalingsafpraak te maken. Bij het ontwerpen van de nudges werd er rekening mee gehouden dat veel ontvangers waarschijnlijk te maken hadden met financiële problemen waardoor ze de schuld niet meteen in een keer konden aflossen. Een evaluatie die de oude brief zonder nudges met de brief met de nudges vergeleek, toonde aan dat de nudges effectief waren. De kans dat de ontvangers een betalingsafpraak maakten was 1.27 keer groter met de nudging brief dan met de oude brief. Dit hoofdstuk toont dus aan dat de aanpak van BIG'R effectief kan zijn en dat je om schuldbetalingen te ontvangen je niet altijd direct de betalingen zelf moet stimuleren maar je je soms kan richten op het stimuleren van het contact opnemen voor een betalingsafpraak.

In **Hoofdstuk 3** werden nudges toegepast om naleving van een veiligheidsmaatregel door medewerkers van de gemeente Rotterdam te bevorderen. Volgens die maatregel waren medewerkers van de gemeente verplicht om een keycord om hun nek te dragen op bepaalde afdelingen zodat herkenbaar was dat zij daar hoorden en zodat zij door anderen aangesproken en naar hun bevoegdheid gevraagd konden worden. Echter droeg ongeveer de helft van de medewerkers geen keycord. BIG'R heeft daarom twee nudges ontwikkeld en in de praktijk getest. De nudges hadden als doel om de medewerkers aan hun keycord te herinneren en om hun motivatie te verhogen deze te dragen. Beide nudges bleken echter ineffectief en leidden zelfs tot negatieve reacties bij sommige medewerkers. Achteraf is daarom een enquête afgenomen onder de medewerkers. Hieruit bleek dat veel medewerkers de nudges als negatief hadden ervaren omdat de medewerkers weinig begrip hadden voor de veiligheidsmaatregel of het er niet mee eens waren. Dit hoofdstuk laat dus zien dat men rekening moet houden met de context en eventuele weerstand bij het ontwikkelen van effectieve nudges.

In **Hoofdstuk 4** gaat het over nudges die gecombineerd zijn met een andere beleidsmaatregel en op die manier beter rekening hielden met de context. De nudges hadden als doel om een huis-aan-huis voorlichtingscampagne, ter vermindering van afval op straat, effectiever te maken. In sommige wijken van Rotterdam plaatsten bewoners hun afval regelmatig niet in een afvalcontainer maar juist ernaast omdat het afval bijvoorbeeld te groot was voor de opening van de container. Dit soort afval wordt ook wel naastplaatsingen genoemd. De nudges bestonden uit stickers die publieksvoorlichters na hun voorlichtingsbezoek uitdeelden aan de bewoners. Door de sticker op of naast hun voordeur te plakken "verplichtten" de bewoners zichzelf als het ware tot het schoonhouden van hun straat. Tegelijkertijd werd op die manier een positieve sociale norm zichtbaar gemaakt die ook werd versterkt door extra borden naast de containers te zetten. Deze nudges bleken erg effectief en het aantal dagen dat er naastplaatsingen werden gevonden nam met twee derde af. Dit hoofdstuk laat de potentie van



nudges zien wanneer ze gecombineerd worden met en geïntegreerd worden in bestaand beleid (zie ook de Ridder et al, *forthcoming*).

**Hoofdstuk 5** is een voorbeeld van het maken van beleid met gedragsinzichten tijdens de formuleringsfase van beleid. Het gaat dus ook niet om een interventie maar om meer algemene beleidsstrategieën. Specifiek gaat het over het gebruik van vuurwerk door voetbalsupporters tijdens wedstrijden in de Kuip, het voetbalstadion van Feyenoord Rotterdam. Middels interviews met supporters werden hun beweegredenen achterhaald. Deze inzichten werden naast de visies van verschillende professionals, die in hun werk met het vuurwerk te maken kregen, gelegd. Uit de interviews bleek dat het verhogen van de sfeer in het stadion en steun bieden aan de spelers op het veld, hoorden bij de belangrijkste beweegredenen. Vanuit een gedragsperspectief bleek dat het in gesprek gaan van supporters en professionals en het zoeken van compromissen de meest geschikte strategie zou zijn. Dit hoofdstuk laat de potentie van een gedragsaanpak zien om beleidsstrategieën te ontwikkelen, en toont aan wat de bijdrage van gedragsinzichten kan zijn als gedragsteams zich niet alleen maar richten op het uitvoeren van experimenten.

In **Hoofdstuk 6** wordt BIG'R in zijn geheel geëvalueerd. Hiertoe werden de *inputs* (bijvoorbeeld personeel, budget), de activiteiten (bijvoorbeeld projecten, trainingen), de *outputs* (bijvoorbeeld toegepaste beleidsadviezen) en de beleidsuitkomsten (bijvoorbeeld minder afval op straat) van BIG'R systematisch beschreven en onderzocht (Saunders et al., 2005). Daarnaast werd ook rekening gehouden met de context waarin de groep heeft gewerkt. Aan het onderzoek namen onder andere leden van BIG'R, opdrachtgevers en cursisten deel door middel van vragenlijsten en/of interviews. De resultaten zijn uiteindelijk verwerkt in negen stellingen die beschrijven hoe je het werken met gedragsinzichten kan vormgeven in een overheidsorganisatie (tabel S.1).

Hoofdstuk 6 is naar mijn weten de eerste wetenschappelijke evaluatie van een geheel gedragsteam. Het laat daardoor ook de verhouding van inputs tot outputs en beleidsuitkomsten zien. Evaluaties in het verleden concentreerden zich vooral op de outputs van gedragsteams, met name hun interventies (bijv. DellaVigna & Linos, 2020), maar niet gedragsteams an sich en hoe ze kunnen worden ingebed in een organisatie.

## Sectie 3

Sectie 3 bestaat uit twee hoofdstukken die mogelijke manieren aantonen om toekomstig beleid te maken met gedragsinzichten.

In **Hoofdstuk 7** werd aan de hand van een vragenlijst onderzocht hoe aannames van beleidsmedewerkers over gedragsverandering hun beleidskeuzes kunnen beïnvloeden. Uit de resultaten bleek dat beleidsmedewerkers die ervan uitgaan dat gedrag onveranderbaar of moeilijk is om aan te passen een sterkere voorkeur hadden voor gedragsverandering via strenge beleidsmaatregelen en voorlichting dan beleidsmedewerkers voor wie gedrag wel veranderbaar was. Een verklaring hiervoor kan zijn dat volgens beleidsmedewerkers meer overheidsinzet nodig is om gedrag te veranderen wanneer ze ervan uitgaan dat mensen hun gedrag niet zelfstandig kunnen aanpassen.

**Tabel S.1****Stellingen**

Stelling	
1	Beleidsprofessionals die werkzaam zijn binnen gedragsteams hebben competenties voor het uitvoeren of managen van onderzoek en kennis over gedragswetenschap nodig.
2	Gedragsteams moeten hun aanpak aanpassen (bijvoorbeeld onderzoeksmethoden, eigen rol) om onder verschillende administratieve omstandigheden en met verschillende doelgedragingen te kunnen functioneren.
3	Stabiliteit van het team en/of een goede overdracht aan nieuwe teamleden zijn de sleutel tot de succesvolle afronding van projecten en zijn essentieel om het lerend vermogen en de ontwikkeling van de groep te verbeteren.
4	Om de reikwijdte van gedragsverandering te vergroten, hebben gedragsteams een breed mandaat nodig dat ook ondersteuning voor implementatie omvat.
5	Naast directe vormen van bewijs (bijvoorbeeld veldexperimenten), kunnen gedragsteams ook profiteren van indirecte vormen van bewijs (bijvoorbeeld deskundige autoriteit) om invloed uit te oefenen.
6	Het uitleggen van de mechanismen die ten grondslag liggen aan gedragsoplossingen aan besluitvormers leidt tot betere implementatie.
7	Om betere implementatie te bereiken kunnen gedragsteams investeren in meer gerichte implementatie-inspanningen en profiteren van inzichten uit de implementatiewetenschap.
8	Beleidsprofessionals kunnen bewust een samenwerking aangaan met een gedragsteam om op die manier meer te weten te komen over het belang en het nut van gedragsinzichten voor openbaar beleid.
9	Effectieve grenswerkers (" <i>boundary workers</i> ") zijn belangrijk om het bereik en de uitvoering van interventies te vergroten.

Dit hoofdstuk is een van de eerste studies die de invloed van gedragsfactoren op keuzes van beleidsmedewerkers onderzocht (Capano & Howlett, 2020; Howlett et al., 2020). Daarbij wordt aangetoond dat de voorkeur voor bepaalde beleidsinstrumenten worden beïnvloed door het beeld dat beleidsmedewerkers hebben over gedrag. Beleid maken met gedragsinzichten gaat dus niet alleen over interventies zoals nudges of beleid op zich maar ook over denkwijzen van beleidsmedewerkers. In de toekomst zou je hierop in kunnen spelen op een positieve manier, bijvoorbeeld door ervoor te zorgen dat beleidsmedewerkers bij hun keuzes uitgaan van een realistisch beeld over gedrag.

**Hoofdstuk 8** gaat over verschillende strategieën die verschillende gedragsteams zouden kunnen toepassen. Drie strategische dimensies, die zijn ontworpen op basis van de ervaringen met BIG'R, worden beschreven en uitgelegd. Eén dimensie beschrijft of het gedragsteam vooral wetenschappelijk werkt om nieuwe inzichten op te leveren, of vooral praktijkgericht met als doel het oplossen van beleidsvraagstukken. Een andere dimensie beschrijft het verschil tussen gedragsteams die als doel hebben om op een efficiënte manier bruikbare interventies te ontwikkelen en gedragsteams die vooral

voor zichzelf en anderen willen leren hoe je beleid en gedraginsichten kan combineren. De laatste dimensie beschrijft de coördinatie van leden van een gedragsteam, waarbij gekozen kan worden voor een duidelijke hiërarchie of voor een netwerk aan teamleden die zonder hiërarchische structuur samenwerken.

Dit hoofdstuk omvat het eerste onderzoek naar verschillende strategieën van gedragsteams en toont op die manier ook de verscheidenheid aan van hoe je beleid zou kunnen maken met gedraginsichten. Mensen in de praktijk kunnen daarnaast gebruik maken van de dimensies als een reflectiehulp om bewuste keuzes te maken over het type onderzoek dat en uitkomsten die zij willen bereiken met hun eigen gedragsteam.

## Discussie

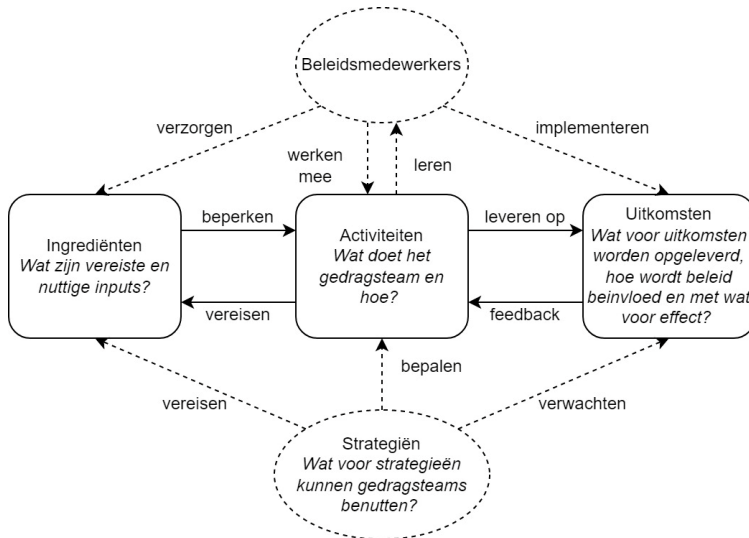
In het laatste hoofdstuk wordt de onderzoeksvraag (hoe kan je door middel van gedragsteams beleid maken met gedraginsichten?) beantwoord door de resultaten van de voorafgaande secties en hoofdstukken te combineren in een theoretisch model over de werkwijze van gedragsteams (afbeelding S.1). De kern van het model bestaat uit de ingrediënten die een gedragsteam nodig heeft, de activiteiten van het team en de uitkomsten en effecten die het team behaalt.

De ingrediënten zijn noodzakelijk voor het uitvoeren van de activiteiten van een gedragsteam. Voor BIG'R zijn de ingrediënten in hoofdstuk 1 beschreven en in hoofdstuk 6 geëvalueerd. De hoofdactiviteit van veel gedragsteams bestaat, zoals voor BIG'R, uit het onderzoeken van concrete beleidsvraagstukken vanuit een gedragsperspectief en oplossingen bedenken om gedrag te veranderen. Daarnaast trainen veel gedragsteams, inclusief BIG'R, ook beleidsmedewerkers zodat zij zelf aan de slag kunnen gaan met het toepassen van gedraginsichten (bijvoorbeeld Baggio et al., 2021). De uitkomsten die gedragsteams opleveren zoals beleidsadviezen worden idealiter door beleidsmedewerkers buiten het gedragsteam geïmplementeerd met positieve gevolgen. In de praktijk van BIG'R werden echter niet alle adviezen geïmplementeerd en bleken sommige interventies van BIG'R niet effectief (bijvoorbeeld hoofdstuk 3). Daarom heeft BIG'R in de loop van de tijd steeds meer aandacht besteed aan implementatie.

Daarnaast houdt het model rekening met de belangrijke rol van externe beleidsmedewerkers om benodigde ingrediënten beschikbaar te stellen voor gedragsteams (bijvoorbeeld financiën), om de activiteiten van de teams in samenwerking uit te voeren (bijvoorbeeld informatie doorgeven, eisen voor interventies beschrijven) en om adviezen van de teams te implementeren en instrumenten toe te passen. Beleidsmedewerkers krijgen in de interactie met gedragsteams ook nieuwe inzichten over gedrag en leren over het belang van gedrag voor effectiever beleid. Daarnaast houdt het model er ook rekening mee dat verschillende gedragsteams verschillende strategieën kunnen toepassen en hun ingrediënten, activiteiten en hun uitkomsten daardoor worden beïnvloed. Daarom zijn externe beleidsmedewerkers en strategieën van gedragsteams allebei onderdeel van het model. Naar mijn weten is dit het eerste model over de werkwijze van gedragsteams dat in de wetenschappelijke literatuur is beschreven.

## Afbeelding S.1

### Theoretisch model van gedragsteams



Daarnaast bespreek ik methodologische aspecten die van toepassing zijn op het onderzoek. De werkwijze van BIG'R en het onderzoek in dit proefschrift zijn gebaseerd op een pragmatische aanpak voor onderzoek (Morgan, 2007; 2014). Een pragmatische aanpak is in essentie niet geïnteresseerd in abstracte discussies over de aard van kennis en met welke methodologieën die kennis opgedaan kan worden; een pragmatische aanpak heeft als doel om betekenisvol onderzoek uit te voeren dat keuzes en acties van mensen in de praktijk kan beïnvloeden. Het laat dan ook toe dat onderzoekers verschillende manieren van onderzoek (bijvoorbeeld kwalitatief en kwantitatief onderzoek) combineren en voor elk onderzoeksvraag de beste manier kiezen. Het grootste voordeel van een pragmatische aanpak voor gedragsteams is dat zij daardoor in staat zijn om meer uiteenlopende beleidsvraagstukken op te pakken dan wanneer teams zich beperken tot één manier van onderzoek. In het verleden hebben gedragsteams zich vaak beperkt tot het uitvoeren van gerandomiseerde onderzoeken met controlegroep (bijvoorbeeld Jones et al., 2021). Hoofdstuk 5 van dit proefstuk is een voorbeeld van het tegenovergestelde daarvan en illustreert dat je ook door middel van interviews een belangrijke bijdrage kan leveren aan beleid.

Daarnaast wordt nog ingegaan op co-creatie als een essentieel element van zowel de werkwijze van BIG'R en sommige onderzoeken van dit proefschrift, met name de hoofdstukken 2 tot en met 5 die voorbeelden uit de praktijk van BIG'R weergeven. Het onderzoek is dus tot stand gekomen en uitgevoerd in samenwerking met beleidsmedewerkers, bijvoorbeeld om de relevantie van beleidsvraagstukken die werden aangedragen aan BIG'R te beoordelen. Co-creatie wordt vaak gezien als een geschikte manier om het gat tussen wetenschappelijke kennis en de praktijk te verkleinen (Marshall et al., 2014; Walter et al., 2003). Daarnaast draagt co-creatie bij aan de ontwikkeling van

beleidsadviezen die rekening houden met lokale omstandigheden en niet-academische kennis uit de praktijk.

Afsluitend worden op basis van het onderzoek implicaties voor de praktijk toegelicht en suggesties gedaan voor de doorontwikkeling voor het maken van beleid met gedragsinzichten bij de Gemeente Rotterdam. De eerste implicatie voor de praktijk is dat gedragsteams het beste kunnen worden ingezet om beleidsvraagstukken op te pakken die voldoende relevant zijn om de extra inzet van het team te rechtvaardigen maar anderzijds niet zo belangrijk zijn zodat falen geen optie is. Want het onderzoek toont aan dat oplossingen die door gedragsteams worden ontwikkeld effectief kunnen zijn maar dat er geen garantie is voor effectiviteit. De tweede implicatie is dat gedragsteams zich het beste kunnen bezighouden met beleidsvraagstukken die met normale beleidsaanpakken niet voldoende opgelost kunnen worden. Gedragsinzichten worden namelijk veelvuldig verspreid, en inmiddels weten ook bij overheden steeds meer mensen hoe ze rekening kunnen houden met gedrag. De expertise van gedragsteams is daarom vooral te zien als een aanvulling daarop voor de ingewikkeldste gedragsvraagstukken. Ten derde moeten gedragsteams onderdeel zijn van een strategie voor het maken van beleid met gedragsinzichten om bijvoorbeeld de doelstellingen voor het team helder te kunnen omschrijven. De suggesties voor de Gemeente Rotterdam zijn ten eerste gericht op het idee om gedragskennis via een netwerk van beleidsmedewerkers verder te verdiepen en te verspreiden, ten tweede op het implementeren van de opbrengsten van BIG'R en ten derde op het benutten van de kennis van voormalige BIG'R medewerkers.

Als laatste worden in dit hoofdstuk drie veelbelovende onderwerpen voor toekomstig onderzoek aangedragen. Het eerste onderwerp is verder onderzoek naar de werkwijze van gedragsteams om beter te weten te komen wat eraan kan bijdragen dat zij succesvol zijn. Het tweede onderwerp betreft onderzoek rondom de invloed van gedragsteams op de beleidscontext. Het laatste onderwerp is het doorgaan met de ontwikkeling van effectieve gedragsinterventies die bij kunnen dragen aan effectief beleid.



## References

## A

- Afif, Z., Islan, W. W., Calvo-Gonzalez, O., & Dalton, A. G. (2018). *Behavioral science around the world: Profiles of 10 countries* (Report No. 132610). World Bank Group.  
<http://documents.worldbank.org/curated/en/710771543609067500/Behavioral-Science-Around-the-World-Profiles-of-10-Countries>
- Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3), 715–753. <https://doi.org/10.1162/003355300554881>
- Al-Ubaydli, O., Lee, M. S., List, J. A., Mackevicius, C. L., & 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>
- Ames, P., & Hiscox, M. (2016). *Guide to developing behavioural interventions for randomised controlled trials*. Australia Government. <https://www.dpc.nsw.gov.au/assets/dpc-nsw-gov-au/publications/Behavioural-Insights-Unit/7091d27d37/BETA-Guide-to-developing-behavioural-interventions-for-randomised-controlled-trials.pdf>
- Amos, C., & Paswan, A. (2009). Getting past the trash bin: Attribution about envelope message, envelope characteristics, and intention to open direct mail. *Journal of Marketing Communications*, 15(4), 247–265. <https://doi.org/10.1080/13527260802332053>
- Anderson, P. W. (1972). More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047), 393–396.  
<https://doi.org/10.1126/science.177.4047.393>
- Anderson, J. E. (1984). *Public policy-making: An introduction* (3rd ed.). Holt, Rinehart and Winston.
- Anglin, G. J., Vaez, H., & Cunningham, K. L. (2004). Visual representations and learning: The role of static and animated graphics. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 865–916). Lawrence Erlbaum Associates Publishers.
- Ansell, C., Sørensen, E., & Torfing, J. (2017). Improving policy implementation through collaborative policymaking. *Policy & Politics*, 45(3), 467–486.  
<https://doi.org/10.1332/030557317x14972799760260>
- Antinyan, A., & Asatryan, Z. (2020). *Nudging for tax compliance: A meta-analysis* (CESifo Working Paper No. 8500). Center for Economic Studies and Ifo Institute.  
<http://hdl.handle.net/10419/223572>
- Ariely, D. (2008). *Predictably irrational*. HarperCollins.
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40(4), 471–499.  
<https://doi.org/10.1348/014466601164939>
- Atkins, L. & Wallace, S. (2012). Insider research. In *Qualitative research in education* (pp. 47–64). SAGE Publications. <https://www.doi.org/10.4135/9781473957602>



Auditteam Voetbal en Veiligheid. (2017). *Vuurwerk tijdens en rond voetbalwedstrijden: No pyro, no party? Aanbevelingen op de aanpak van vuurwerk in het betaald voetbal te verbeteren*. Ministerie van Justitie en Veiligheid

Árvai, J., & Gregory, R. (2021). Beyond choice architecture: A building code for structuring climate risk management decisions. *Behavioural Public Policy*, 5(4), 556–575.  
<http://doi.org/10.1017/bpp.2020.37>

## B

Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E. A., & Nelson, L. D. (2012). Commitment and behavior change: Evidence from the field. *Journal of Consumer Research*, 39(5), 1070–1084.  
<https://doi.org/10.1086/667226>

Baekgaard, M., Belle, N., Serritzlew, S., Sicilia, M., & Steccolini, I. (2019). Performance information in politics: How framing, format, and rhetoric matter to politicians' preferences. *Journal of Behavioral Public Administration*, 2(2). <https://doi.org/10.30636/jbpa.22.67>

Baggio, M., Ciriolo, E., Marandola, G., & van Bavel, R. (2021). The evolution of behaviourally informed policy-making in the EU. *Journal of European Public Policy*, 28(5), 658–676.  
<https://doi.org/10.1080/13501763.2021.1912145>

Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources model: State of the art. *Journal of Managerial Psychology*, 22(3), 309–328. <https://doi.org/10.1108/02683940710733115>

Ball, S. (forthcoming). Translating behavioral public policy into practice: Interpretations and traditions. *Critical Policy Studies*. <https://doi.org/10.1080/19460171.2021.1973525>

Ball, S., & Feitsma, J. (2020). The boundaries of behavioural insights: Observations from two ethnographic studies. *Evidence & Policy: A Journal of Research, Debate and Practice*, 16(4), 559–577. <https://doi.org/10.1332/174426419X15643724702722>

Ball, S., & Head, B. W. (2021). Behavioural insights teams in practice: Nudge missions and methods on trial. *Policy & Politics*, 49(1), 105–120. <https://doi.org/10.1332/030557320X15840777045205>

Ball, S., Hiscox, M., & Oliver, T. (2017). Starting a behavioural insights team: Three lessons from the Behavioural Economics Team of the Australian Government. *Journal of Behavioral Economics for Policy*, 1(1), 21–26.

Barasz, K., John, L. K., Keenan, E. A., & Norton, M. I. (2017). Pseudo-set framing. *Journal of Experimental Psychology: General*, 146(10), 1460–1477. <https://doi.org/10.1037/xge0000337>

Barnett, C., & Camfield, L. (2016). Ethics in evaluation. *Journal of Development Effectiveness*, 8(4), 528–34. <https://doi.org/10.1080/19439342.2016.1244554>

Beaulieu, M., Breton, M., & Brousselle, A. (2018). Conceptualizing 20 years of engaged scholarship: A scoping review. *PloS one*, 13(2), e0193201. <https://doi.org/10.1371/journal.pone.0193201>

Behavioural Insights Team. (2012). *Applying behavioural insights to reduce fraud, debt and error*.

Behavioural Insights Team. (2019). *Annual Report 2017-18*.

- 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>
- Ben-Porat, A. (2010). Football fandom: A bounded identification. *Soccer & Society*, 11(3), 277–290. <https://doi.org/10.1080/14660971003619594>
- Bensley, D. A., & Lilienfeld, S. O. (2017). Psychological misconceptions: Recent scientific advances and unresolved issues. *Current Directions in Psychological Science*, 26(4), 377–382. <https://doi.org/10.1177/0963721417699026>
- Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics*, 127(3), 1205–1242. <https://doi.org/10.1093/qje/qjs017>
- Birukou A., Blanzieri E., Giorgini P., & Giunchiglia, F. (2013). A formal definition of culture. In K. Sycara, M. Gelfand, & A. Abbe (Eds.), *Models for intercultural collaboration and negotiation* (pp. 1–26). Springer. [https://doi.org/10.1007/978-94-007-5574-1\\_1](https://doi.org/10.1007/978-94-007-5574-1_1)
- Boeije, H. (2002). A Purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality and Quantity*, 36(4), 391–409. <https://doi.org/10.1023/A:1020909529486>
- Bolton, A., & Newell, B. R. (2017). Applying behavioural science to government policy: Finding the goldilocks zone. *Journal of Behavioral Economics for Policy*, 1(1), 9–14.
- Boud, D. (2001). Using journal writing to enhance reflective practice. *New Directions for Adult and Continuing Education*, 2001(90), 9–18. <https://doi.org/10.1002/ace.16>
- Boyer, E. (1996). The scholarship of engagement. *Journal of Public Service and Outreach*, 1(1), 11–20.
- Brannick, T., & Coghlan, D. (2007). In defense of being “native”: The case for insider academic research. *Organizational Research Methods*, 10(1), 59–74. <https://doi.org/10.1177/1094428106289253>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brechtbühl, A., Schumacher Dimech, A., Schmid, O. N., & Seiler, R. (2017). Escalation vs. non-escalation of fan violence in football? Narratives from ultra fans, police officers and security employees. *Sport in Society*, 20(7), 861–879. <https://doi.org/10.1080/17430437.2016.1221932>
- Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.
- Brehm, J. W., & Brehm, S. S. (1981). *Psychological reactance: A theory of freedom and control*. Academic Press.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32(7), 513–531. <https://doi.org/10.1037/0003-066X.32.7.513>
- Brown, G., & Raymond, C. (2006). The relationship between place attachment and landscape values: Toward mapping place attachment. *Applied Geography*, 27(2), 89–111. <https://doi.org/10.1016/j.apgeog.2006.11.002>

- Brown, S., Taylor, K., & Price, S. W. (2005). Debt and distress: Evaluating the psychological cost of credit. *Journal of Economic Psychology*, 26(5), 642–663.  
<https://doi.org/10.1016/j.joep.2005.01.002>
- Brownson, R. C., Royer, C., Ewing, R., & McBride, T. D. (2006). Researchers and policymakers: Travelers in parallel universes. *American Journal of Preventive Medicine*, 30(2), 164–172.  
<https://doi.org/10.1016/j.amepre.2005.10.004>
- Budim A. (2018). *Sport matters: Hooliganism and corruption in football* [Unpublished bachelor's dissertation]. Josip Juraj Strossmayer University of Osijek.
- Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*, 34(3), 523–548. <https://doi.org/10.2307/25750690>
- Burns, T., & Stalker, G. M. (1994). *The management of innovation* (Rev. ed.). Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780198288787.001.0001>

## C

- Cadario, R., & Chandon, P. (2020). Which healthy eating nudges work best? A meta-analysis of field experiments. *Marketing Science*, 39(3), 465–486. <https://doi.org/10.1287/mksc.2018.1128>
- Cairney, P., & Oliver, K. (2020). How should academics engage in policymaking to achieve impact? *Political Studies Review*, 18(2), 228–244. <https://doi.org/10.1177/1478929918807714>
- 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>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmeld, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Capano, G., & Howlett, M. (2020). The knowns and unknowns of policy instrument analysis: Policy tools and the current research agenda on policy mixes. *SAGE Open*, 10(1).  
<https://doi.org/10.1177/2158244019900568>
- Capano, G., & Lippi, A. (2017). How policy instruments are chosen: patterns of decision makers' choices. *Policy Sciences*, 17(2), 269–293. <https://doi.org/10.1007/s11077-016-9267-8>
- Caplan, N. (1979). The two-communities theory and knowledge utilization. *American Behavioral Scientist*, 22(3), 459–470. <https://doi.org/10.1177/000276427902200308>
- Caris, M. G., Labuschagne, H. A., Dekker, M., Kramer, M., van Agtmael, M. A., & Vandenbroucke-Grauls, C. (2018). Nudging to improve hand hygiene. *The Journal of Hospital Infection*, 98(4), 352–358. <https://doi.org/10.1016/j.jhin.2017.09.023>
- Cartwright, N. (2010). What are randomised controlled trials good for? *Philosophical Studies*, 147(1), 59. <https://doi.org/10.1007/s11098-009-9450-2>

- Casakin, H., & Reizer, A. (2017). Place attachment, residential satisfaction, and life satisfaction: Traditional and renewed kibbutz. *Journal of Human Behavior in the Social Environment*, 27(7), 639–655. <https://doi.org/10.1080/10911359.2017.1317313>
- Centraal Bureau voor de Statistiek. (2018). *Armoede en sociale uitsluiting*.
- Chan, M., Woon, I., & Kankanhalli, A. (2005). Perceptions of information security in the workplace: Linking information security climate to compliant behavior. *Journal of Information Privacy and Security*, 1(3), 18–41. <https://doi.org/10.1080/15536548.2005.10855772>
- Chapman, G., Milkman, K. L., Rand, D., Rogers, T., & Thaler, R. H. (2021). Nudges and choice architecture in organizations: New frontiers. *Organizational Behavior and Human Decision Processes*, 163, 1–3. <https://doi.org/10.1016/j.obhdp.2020.04.004>
- Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5), 1–33. <https://doi.org/10.1257/aer.p20151108>
- Chiu, C.-Y., Morris, M. W., Hong, Y.-Y., & Menon, T. (2000). Motivated cultural cognition: The impact of implicit cultural theories on dispositional attribution varies as a function of need for closure. *Journal of Personality and Social Psychology*, 78(2), 247–259. <https://doi.org/10.1037/0022-3514.78.2.247>
- Choi, B. C., Pang, T., Lin, V., Puska, P., Sherman, G., Goddard, M., Ackland, M. J., Sainsbury, P., Stachenko, S., Morrison, H., & Clotey, C. (2005). Can scientists and policy makers work together? *Journal of Epidemiology and Community Health*, 59(8), 632–637. <https://doi.org/10.1136/jech.2004.031765>
- Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, 94(5), 1103–1127. <https://doi.org/10.1037/a0016172>
- Chung, A. H., & Slater, M. D. (2013). Reducing stigma and out-group distinctions through perspective-taking in narratives. *Journal of Communication*, 63(5), 894–911. <https://doi.org/10.1111/jcom.12050>
- Cialdini, R. B. (1984). *Influence: The psychology of persuasion*. William Morrow and Company.
- Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika*, 72(2), 263–268. <https://doi.org/10.1007/s11336-006-1560-6>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55(1), 591–621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015–1026. <https://doi.org/10.1037/0022-3514.58.6.1015>
- Clayton, M., Liñares-Zegarra, J., & Wilson, J. (2015). Does debt affect health? Cross country evidence on the debt-health nexus. *Social Science & Medicine*, 130, 51–58. <https://doi.org/10.1016/j.socscimed.2015.02.002>
- Coghlan, D. (1993). Learning from emotions through journaling. *Journal of Management Education*, 17(1), 90–94. <https://doi.org/10.1177/105256299301700107>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Colquitt, J. A., Greenberg, J., & Scott, B. A. (2005). Organizational justice: Where do we stand? In J. Greenberg & J. A. Colquitt (Eds.), *Handbook of organizational justice* (pp. 589–619). Lawrence Erlbaum Associates.
- Congdon, W. J., & Shankar, M. (2015). The White House Social & Behavioral Sciences Team: Lessons learned from year one. *Behavioral Science & Policy*, 1(2), 77–86. <https://doi.org/10.1353/bsp.2015.0010>
- Congdon, W. J., & Shankar, M. (2018). The role of behavioral economics in evidence-based policymaking. *The ANNALS of the American Academy of Political and Social Science*, 678(1), 81–92. <https://doi.org/10.1177/0002716218766268>
- Cotterill, S., John, P., & Johnston, M. (2021). How can better monitoring, reporting and evaluation standards advance behavioural public policy? *Policy & Politics*, 49(1), 161–179. <https://doi.org/10.1332/030557320x15955052119363>
- Coughlan, P., Suri, J. F., & Canales, K. (2007). Prototypes as (design) tools for behavioral and organizational change: A design-based approach to help organizations change work behaviors. *The Journal of Applied Behavioral Science*, 43(1), 122–134. <https://doi.org/10.1177/0021886306297722>
- Curchin, K. (2016). Using behavioural insights to argue for a stronger social safety net: Beyond libertarian paternalism. *Journal of Social Policy*, 46(2), 231–249. <https://doi.org/10.1017/s0047279416000672>

## D

- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexanders, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4, 50. <https://doi.org/10.1186/1748-5908-4-50>
- Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>
- de Bruijn, E.-J., & Antonides, G. (forthcoming). Poverty and economic decision making: A review of scarcity theory. *Theory and Decision*. <https://doi.org/10.1007/s11238-021-09802-7>
- De Cremer, D., Ouden, N. d., & Tyler, T. R. (2005). Managing cooperation via procedural fairness: The mediating influence of self-other merging. *Journal of Economic Psychology*, 26(3), 393–406. <https://doi.org/10.1016/j.joep.2004.12.004>
- De Cremer, D., & Tyler, T. R. (2007). The effects of trust in authority and procedural fairness on cooperation. *Journal of Applied Psychology*, 92(3), 639–649. <https://doi.org/10.1037/0021-9010.92.3.639>

- Dedehayir, O., & Steinert, M. (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, 108, 28–41. <https://doi.org/10.1016/j.techfore.2016.04.005>
- DeJoy, D. M. (2005). Behavior change versus culture change: Divergent approaches to managing workplace safety. *Safety Science*, 43(2), 105–129. <https://doi.org/10.1016/j.ssci.2005.02.001>
- de Kort, Y. A. W., McCalley, L. T., & Midden, C. J. H. (2008). Persuasive trash cans: Activation of littering norms by design. *Environment and Behavior*, 40(6), 870–891. <https://doi.org/10.1177/0013916507311035>
- DellaVigna, S., & Linos, E. (2020). *RCTs to scale: Comprehensive evidence from two nudge units* (Working Paper No. 27594). National Bureau of Economic Research. <https://doi.org/10.3386/w27594>
- 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>
- Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*. Aldine Publishing Company.
- de Ridder, D., Feitsma, J., van den Hoven, M., Kroese, F., Schillemans, T., Verweij, M., Venema, T., Vugts, A., & de Vet, E. (forthcoming). Simple nudges that are not so easy. *Behavioural Public Policy*. <https://doi.org/10.1017/bpp.2020.36>
- de Wijk, R. A., Maaskant, A. J., Polet, I. A., Holthuysen, N. T. E., van Kleef, E., & Vingerhoeds, M. H. (2016). An in-store experiment on the effect of accessibility on sales of wholegrain and white bread in supermarkets. *PLoS ONE*, 11(3), e0151915. <https://doi.org/10.1371/journal.pone.0151915>
- Diener, E., & Wallbom, M. (1976). Effects of self-awareness on antinormative behavior. *Journal of Research in Personality*, 10(1), 107–111. [https://doi.org/10.1016/0092-6566\(76\)90088-X](https://doi.org/10.1016/0092-6566(76)90088-X)
- Dimant, E. (2019). Contagion of pro- and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73, 66–88. <https://doi.org/10.1016/j.joep.2019.04.009>
- Dionísio, P., Leal, C., & Moutinho, L. (2008). Fandom affiliation and tribal behaviour: A sports marketing application. *Qualitative Market Research*, 11(1), 17–39. <https://doi.org/10.1108/13522750810845531>
- Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, 33(1), 264–277. <https://doi.org/10.1016/j.joep.2011.10.009>
- Dudley, S. E., & Xie, Z. (2020). Designing a choice architecture for regulators. *Public Administration Review*, 80(1), 151–156. <https://doi.org/10.1111/puar.13112>
- Dunleavy, K., Noble, M., & Andrews, H. (2019). The emergence of the publicly engaged research manager. *Research for All*, 3(1), 105–24. <https://doi.org/10.18546/rfa.03.1.09>

- Dur, R., & Vollaard, B. (2015). The power of a bad example: A field experiment in household garbage disposal. *Environment and Behavior*, 47(9), 970–1000.  
<https://doi.org/10.1177/0013916514535085>
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327–350. <https://doi.org/10.1007/s10464-008-9165-0>
- Dye, T. R. (1972). *Understanding public policy*. Prentice-Hall.

## E

- Einfeld, C. (2019). Nudge and evidence based policy: fertile ground. *Evidence & Policy: A Journal of Research, Debate and Practice*, 15(4), 509–524.  
<https://doi.org/10.1332/174426418X15314036559759>
- Einfeld, C., & Blomkamp, E. (forthcoming). Nudge and co-design: Complementary or contradictory approaches to policy innovation? *Policy Studies*.  
<https://doi.org/10.1080/01442872.2021.1879036>
- Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58(4), 637–642.  
<https://doi.org/10.1007/s00038-012-0416-3>
- Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes* (2nd ed.). The University of Chicago Press.
- Etzkowitz, H. & Leydesdorff, L. (2000). The dynamics of innovation: From national systems and “mode 2” to a triple helix of university–industry–government relations. *Research Policy*, 29(2), 109–123.  
[https://doi.org/10.1016/S0048-7333\(99\)00055-4](https://doi.org/10.1016/S0048-7333(99)00055-4)
- 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>
- Ewert, B., & Loer, K. (2021). Advancing behavioural public policies: In pursuit of a more comprehensive concept. *Policy & Politics*, 49(1), 25–47.  
<https://doi.org/10.1332/030557320x15907721287475>
- Exec. Order No. 13707, 3 C.F.R. 371 (2015). <https://www.govinfo.gov/content/pkg/CFR-2016-title3-vol1/pdf/CFR-2016-title3-vol1-eo13707.pdf>

## F

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Feitsma, J. (2018). The behavioural state: Critical observations on technocracy and psychocracy. *Policy Sciences*, 51(3), 387–410. <https://doi.org/10.1007/s11077-018-9325-5>

- Feitsma, J. (2019a). Brokering behaviour change: The work of behavioural insights experts in government. *Policy & Politics*, 47(1), 37-56.  
<https://doi.org/10.1332/030557318X15174915040678>
- Feitsma, J. (2019b). *Inside the behavioural state*. Eleven International Publishing.
- 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>
- Feitsma, J., & Schillemans, T. (2019). Behaviour experts in government: From newcomers to professionals? In H. Straßheim, & S. Beck (Eds.), *Handbook of behavioural change and public policy* (pp. 122–137). Edward Elgar Publishing. <https://doi.org/10.4337/9781785367854.00015>
- Feitsma, J., & Whitehead, M. (2021). *Dynamics of behavioural expertise under COVID-19*. Advance.  
<https://doi.org/10.31124/advance.13725601.v2>
- Feitsma, J., & Whitehead, M. (forthcoming). Bounded interdisciplinarity: Critical interdisciplinary perspectives on context and evidence in behavioural public policies. *Behavioural Public Policy*.  
<http://doi.org/10.1017/bpp.2019.30>
- Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1), 541–556.  
<https://doi.org/10.1257/aer.100.1.541>
- Fitch, C., Hamilton, S., Bassett, P., & Davey, R. (2011). The relationship between personal debt and mental health: A systematic review. *Mental Health Review Journal*, 16(4), 153–166.  
<https://doi.org/10.1108/13619321111202313>
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). National Implementation Research Network. <http://www.fpg.unc.edu/nirn/resources/publications/Monograph/>
- Flinders, M., Wood, M., & Cunningham, M. (2016). The politics of co-production: Risks, limits and pollution. *Evidence & Policy: A Journal of Research, Debate and Practice*, 12(2), 261–279.  
<https://doi.org/10.1332/174426415X14412037949967>
- Forrest, R., & Kearns, A. (2001). Social cohesion, social capital and the neighbourhood. *Urban Studies*, 38(12), 2125–2143. <https://doi.org/10.1080/00420980120087081>
- Frechtling, J. A. (2007). *Logic modeling methods in program evaluation*. Jossey-Bass.
- French, J. (2011). Why nudging is not enough. *Journal of Social Marketing*, 1(2), 154–162.  
<https://doi.org/10.1108/20426761111141896>
- Furedi, F. (2011, January 20). *Defending moral autonomy against an army of nudgers*. spiked.  
<https://www.spiked-online.com/2011/01/20/defending-moral-autonomy-against-an-army-of-nudgers/>
- Furman, J. (2016). Applying behavioral sciences in the service of four major economic problems. *Behavioral Science & Policy*, 2(2), 1–9. <https://doi.org/10.1353/bsp.2016.0011>



## G

- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- Gennetian, L. A., & Shafir, E. (2015). The persistence of poverty in the context of financial instability: A behavioral perspective. *Journal of Policy Analysis and Management*, 34(4), 904–936. <https://doi.org/10.1002/pam.21854>
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. SAGE Publications.
- Gibbons, F. X., & Wright, R. A. (1983). Self-focused attention and reactions to conflicting standards. *Journal of Research in Personality*, 17(2), 263–273. [https://doi.org/10.1016/0092-6566\(83\)90036-3](https://doi.org/10.1016/0092-6566(83)90036-3)
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3(1), 20–29. <https://doi.org/10.1111/j.1745-6916.2008.00058.x>
- Gil, G. J. (2002). Soccer and kinship in Argentina: The mother's brother and the heritage of identity. *Soccer & Society*, 3(3), 11–25. <https://doi.org/10.1080/714004888>
- Gofen, A., Moseley, A., Thomann, E., & Weaver, R. K. (2021). Behavioural governance in the policy process: Introduction to the special issue. *Journal of European Public Policy*, 28(5), 633–657. <https://doi.org/10.1080/13501763.2021.1912153>
- Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54(7), 493–503. <https://doi.org/10.1037/0003-066X.54.7.493>
- Gonzalez, C., Lerch, J. F., & Lebiere, C. (2010). Instance-based learning in dynamic decision making. *Cognitive Science*, 27(4), 591–635. [https://doi.org/10.1207/s15516709cog2704\\_2](https://doi.org/10.1207/s15516709cog2704_2)
- Gopalan, M., & Pirog, M. A. (2017). Applying behavioral insights in policy analysis: Recent trends in the United States. *Policy Studies Journal*, 45(S1), S82–S114. <https://doi.org/10.1111/psj.12202>
- Gray, B. (2000). Assessing inter-organizational collaboration: Multiple conceptions and multiple methods. In D. Faulkner, & M. de Rond (Eds.), *Cooperative strategy: Economic, business, and organizational issues* (pp. 243–260). Oxford University Press.
- Green, C., & Simmons, R. (2015). The English disease: has football hooliganism been eliminated or just displaced? In P. Rodríguez, S. Késenne, & R. Koning (Eds.), *The economics of competitive sports* (pp. 39–55). Edward Elgar Publishing. <https://doi.org/10.4337/9781783474769.00009>
- Greenberg, A. E., & Hershfield, H. E. (2019). On shifting consumers from high-interest to low-interest debt. *Financial Planning Review*, 2(1), e1035. <https://doi.org/10.1002/cfp2.1035>
- Greenhalgh, T., & Wieringa, S. (2011). Is it time to drop the 'knowledge translation' metaphor? A critical literature review. *Journal of the Royal Society of Medicine*, 104(12), 501–509. <https://doi.org/10.1258/jrsm.2011.110285>
- 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>

- Guo, K. H., Yuan, Y., Archer, N. P., & Connelly, C. E. (2011). Understanding nonmalicious security violations in the workplace: A composite behavior model. *Journal of Management Information Systems*, 28(2), 203–236. <https://doi.org/10.2753/MIS0742-1222280208>
- Guston, D. H. (2001). Boundary organizations in environmental policy and science: An introduction. *Science, Technology, & Human Values*, 26(4), 399–408. <https://doi.org/10.1177/016224390102600401>

## H

- Hall, C. C., Galvez, M. M., & Sederbaum, I. M. (2014). Assumptions about behavior and choice in response to public assistance. *Policy Insights from the Behavioral and Brain Sciences*, 1(1), 137–143. <https://doi.org/10.1177/2372732214550833>
- 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. *Lancet*, 387(10029), 1743–1752. [https://doi.org/10.1016/S0140-6736\(16\)00215-4](https://doi.org/10.1016/S0140-6736(16)00215-4)
- Hallsworth, M., Egan, M., Rutter, J., & McCrae, J. (2018). *Behavioural Government: Using behavioural science to improve how governments make decisions*. Behavioural Insights Team.
- 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>
- Halpern, D., & Sanders, M. (2016). Nudging by government: Progress, impact, & lessons learned. *Behavioral Science & Policy*, 2(2), 52–65. <https://doi.org/10.1353/bsp.2016.0015>
- Hampton, S., & Adams, R. (2018). Behavioural economics vs social practice theory: Perspectives from inside the United Kingdom government. *Energy Research & Social Science*, 46, 214–224. <https://doi.org/10.1016/j.erss.2018.07.023>
- Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, 7(1), 155–174. <http://doi.org/10.1017/S1867299X00005468>
- Hansen, P. G. (2018). What are we forgetting? *Behavioural Public Policy*, 2(2), 190–197. <https://doi.org/10.1017/bpp.2018.13>
- 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>
- Harvey, N. (2020). Behavioral fatigue: Real phenomenon, naïve construct, or policy contrivance? *Frontiers in Psychology*, 11, 589892. <https://doi.org/10.3389/fpsyg.2020.589892>
- Hauser, O. P., Gino, F., & Norton, M. I. (2018). Budging beliefs, nudging behaviour. *Mind & Society*, 17(1–2), 15–26. <https://doi.org/10.1007/s11299-019-00200-9>

- Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *The Journal of Political Philosophy*, 18(1), 123–136. <https://doi.org/10.1111/j.1467-9760.2009.00351.x>
- Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). *Test, learn, adapt: Developing public policy with randomised controlled trials*. Behavioural Insights Team. <https://www.bi.team/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials/>
- Head, B. W. (2010). Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society*, 29(2), 77–94. <https://doi.org/10.1016/j.polsoc.2010.03.001>
- Heidhues, P., & Köszegi, B. (2010). Exploiting naïvete about self-control in the credit market. *American Economic Review*, 100(5), 2279–303. <https://doi.org/10.1257/aer.100.5.2279>
- Hellinga J. (2019). *Derby Days – De twaalfde man*.
- 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>
- Hilbert, L. P., Noordewier, M. K., & van Dijk, W. W. (2022). The prospective associations between financial scarcity and financial avoidance. *Journal of Economic Psychology*, 88, 102459. <https://doi.org/10.1016/j.joep.2021.102459>
- Hofmann, S. G., & Heinrichs, N. (2002). Disentangling self-descriptions and self-evaluations under conditions of high self-focused attention: Effects of mirror exposure. *Personality and Individual Differences*, 32(4), 611–620. [https://doi.org/10.1016/S0191-8869\(01\)00063-0](https://doi.org/10.1016/S0191-8869(01)00063-0)
- Holleman, G. A., Hooge, I., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, 11, 721. <https://doi.org/10.3389/fpsyg.2020.00721>
- Howlett, M. (2018). Matching policy tools and their targets: Beyond nudges and utility maximisation in policy design. *Policy & Politics*, 46(1), 101–124. <https://doi.org/10.1332/030557317x15053060139376>
- Howlett, M. (2019). Behavioural considerations in public policy: Matching policy tools and their targets. In H. Straßheim & S. Beck (Eds.), *Handbook of behavioural change and public policy* (pp. 78–88). Edward Elgar Publishing. <https://doi.org/10.4337/9781785367854.00011>
- Howlett, M., Ramesh, M., & Capano, G. (2020). Policy-makers, policy-takers and policy tools: Dealing with behavioural issues in policy design. *Journal of Comparative Policy Analysis: Research and Practice*, 22(6), 487–497. <https://doi.org/10.1080/13876988.2020.1774367>
- Hu, Q., Dinev, T., Hart, P., & Cooke, D. (2012). Managing employee compliance with information security policies: The critical role of top management and organizational culture. *Decision Sciences*, 43(4), 615–660. <https://doi.org/10.1111/j.1540-5915.2012.00361.x>
- Huitema, D., & Meijerink, S. (2010). Realizing water transitions: The role of policy entrepreneurs in water policy change. *Ecology and Society*, 15(2), 26. <https://doi.org/10.5751/es-03488-150226>

- Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80, 47–58. <https://doi.org/10.1016/j.socec.2019.03.005>
- Hwang, I., Kim, D., Kim, T., & Kim, S. (2017). Why not comply with information security? An empirical approach for the causes of non-compliance. *Online Information Review*, 41(1), 2–18. <https://doi.org/10.1108/OIR-11-2015-0358>

## I

- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86. <http://dx.doi.org/10.1257/jel.47.1.5>

## J

- Jachimowicz, J., Duncan, S., Weber, E., & Johnson, E. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 3(02), 159–186. <https://doi.org/10.1017/bpp.2018.43>
- Jansen, F. (2015). How bulleted lists and enumerations in formatted paragraphs affect recall and evaluation of functional text. *Information Design Journal*, 21(2), 146–162. <https://doi.org/10.1075/idj.21.2.06jan>
- Janssen, M., Bergsma, E., & Cooper, K. (2017). Positieve gedragsbeïnvloeding om inwoners te activeren. Eigen schuld ... of niet? *Sociaal Bestek*, 79(5), 37–39. <https://doi.org/10.1007/s41196-017-0115-y>
- Jasanoff, S. (1990). *The fifth branch: Science advisers as policymakers*. Harvard University Press.
- Jenkins, W. (1978). *Policy analysis: A political and organizational perspective*. St Martin's Press.
- Jensen, N. H., Reuss, L. F., & Rasmussen, S. (2018). Increasing public debt collection with nudging: Results of two natural field experiments. *Scandinavian Journal of Public Administration*, 22(4), 45–64.
- Jewell, R. T., Simmons, R., & Szymanski, S. (2014). Bad for business? The effects of hooliganism on English professional football clubs. *Journal of Sports Economics*, 15(5), 429–450. <https://doi.org/10.1177/1527002514535169>
- John, P. (2013). Experimentation, behaviour change and public policy. *The Political Quarterly*, 84(2), 238–246. <https://doi.org/10.1111/j.1467-923X.2013.12010.x>
- 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. (2019). The international appeal of behavioural public policy: Is nudge an Anglo-American phenomenon? *Journal of Chinese Governance*, 4(2), 144–162. <https://doi.org/10.1080/23812346.2019.1576264>
- John, P., Smith, G., & Stoker, G. (2009). Nudge nudge, think think: Two strategies for changing civic behaviour. *Political Quarterly*, 80(3), 361–370. <https://doi.org/10.1111/j.1467-923X.2009.02001.x>

- John, P., & Stoker, G. (2019). Rethinking the role of experts and expertise in behavioural public policy. *Policy & Politics*, 47(2), 209–226. <https://doi.org/10.1332/030557319x15526371698257>
- Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649), 1338–1339. <https://doi.org/10.1126/science.1091721>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. <https://doi.org/10.3102/0013189X033007014>
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112–133. <https://doi.org/10.1177/1558689806298224>
- Jones, B. D. (2017). Behavioral rationality as a foundation for public policy studies. *Cognitive Systems Research*, 43, 63–75. <https://doi.org/10.1016/j.cogsys.2017.01.003>
- Jones, S., Head, B. & Ferguson, M. (2021). In search of policy innovation: Behavioural insights teams in Australia and New Zealand. *Australian Journal of Public Administration*, 80(3), 435–452. <https://doi.org/10.1111/1467-8500.12478>
- Jones, R., & Whitehead, M. (2018). ‘Politics done like science’: Critical perspectives on psychological governance and the experimental state. *Environment and Planning D: Society and Space*, 36(2), 313–330. <https://doi.org/10.1177/0263775817748330>

## K

- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Katz, R. L. (1955). *Skills of an effective administrator*. Harvard Business Review Press.
- Keizer, K., Lindenberg, S., & Steg, L. (2011). The reversal effect of prohibition signs. *Group Processes & Intergroup Relations*, 14(5), 681–688. <https://doi.org/10.1177/1368430211398505>
- Keller, P. A., Harlam, B., Loewenstein, G., & Volpp, K. G. (2011). Enhanced active choice: A new method to motivate behavior change. *Journal of Consumer Psychology*, 21(4), 376–383. <https://doi.org/10.1016/j.jcps.2011.06.003>
- Kennedy, D. (2013). A contextual analysis of Europe’s ultra football supporters movement. *Soccer & Society*, 14(2), 132–153. <https://doi.org/10.1080/14660970.2013.776464>
- Kilmann, R. H., & Mitroff, I. I. (1979). Problem defining and the consulting/intervention process. *California Management Review*, 21(3), 26–33. <https://doi.org/10.2307/41165304>
- Kimbell, L., & Bailey, J. (2017). Prototyping and the new spirit of policy-making. *CoDesign*, 13(3), 214–226. <https://doi.org/10.1080/15710882.2017.1355003>
- King, D., Greaves, F., Vlaev, I., & Darzi, A. (2013). Approaches based on behavioral economics could help nudge patients and providers toward lower health spending growth. *Health Affairs*, 32(4), 661–668. <https://doi.org/10.1377/hlthaff.2012.1348>
- Knock, E., Whittles, L., Lees, J., Perez Guzman, P., Verity, R., FitzJohn, R., Gaythorpe, K., Imai, N., Hinsley, W., Okell, L., Rosello, A., Kantas, N., Walters, C., Bhatia, S., Watson, O., Whittaker, C., Cattarino, L., Boonyasiri, A., Djaafara, A., . . . Baguelin, M. (2021). Key epidemiological drivers

- and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Science Translational Medicine*, 13(602), eabg4262. <https://doi.org/10.1126/scitranslmed.abg4262>
- KNVB Expertise. (2015). *Fan Onderzoek Eredivisie 2014/15*.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.
- Kolodko, J., Read, D., & Taj, U. (2016). *Using behavioural insights to reduce littering in the UK*. Clean Up Britain.
- Kondrat, M. E. (1992). Reclaiming the practical: Formal and substantive rationality in social work practice. *Social Service Review*, 66(2), 237–255. <http://www.jstor.org/stable/30012164>
- Kondratjeva, O., Roll, S. P., Bufo, S., & Grinstein-Weiss, M. (2021). Using financial tips to guide debt repayment: Experimental evidence from low- and moderate-income tax filers. *Journal of Behavioral and Experimental Economics*, 90(1), 101612. <https://doi.org/10.1016/j.socec.2020.101612>
- Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, 24(1), 65–75. [https://doi.org/10.1016/S0020-7373\(86\)80040-2](https://doi.org/10.1016/S0020-7373(86)80040-2)
- Kossakowski, R., Antonowicz, D., & Jakubowska, H. (2020). The reproduction of hegemonic masculinity in football fandom: An analysis of the performance of Polish ultras. In R. Magrath, J. Cleland, & E. Anderson (Eds.), *The Palgrave handbook of masculinity and sport*. Palgrave Macmillan. [https://doi.org/10.1007/978-3-030-19799-5\\_29](https://doi.org/10.1007/978-3-030-19799-5_29)
- Kosters, M., & Van der Heijden, J. (2015). From mechanism to virtue: Evaluating nudge theory. *Evaluation*, 21(3), 276–291. <https://doi.org/10.1177/1356389015590218>
- Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research*, 30(3), 411–433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). University of Chicago Press. <https://doi.org/10.7208/chicago/9780226458106.001.0001>

## L

- Langley, A., Lindberg, K., Mørk, B. E., Nicolini, D., Raviola, E., & Walter, L. (2019). Boundary work among groups, occupations, and organizations: From cartography to process. *The Academy of Management Annals*, 13(2), 704–736. <https://doi.org/10.5465/annals.2017.0089>
- Lea, S. E. G. (2020). Debt and overindebtedness: Psychological evidence and its policy implications. *Social Issues and Policy Review*, 15(1), 146–179. <https://doi.org/10.1111/sipr.12074>
- Lee, C., & Ma, L. (2020). The role of policy labs in policy experiment and knowledge transfer: A comparison across the UK, Denmark, and Singapore. *Journal of Comparative Policy Analysis: Research and Practice*, 22(4), 281–297. <https://doi.org/10.1080/13876988.2019.1668657>
- Leggett, W. (2014). The politics of behaviour change: Nudge, neoliberalism and the state. *Policy & Politics*, 42(1), 3–19. <https://doi.org/10.1332/030557312X655576>

- Leigh, A. (2018). *Randomistas*. Yale University Press.
- Lepenies, R., Mackay, K., & Quigley, M. (2018). Three challenges for behavioural science and policy: The empirical, the normative and the political. *Behavioural Public Policy*, 2(2), 174–182. <https://doi.org/10.1017/bpp.2018.18>
- Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Technology Research and Development*, 30(4), 195–232. <https://doi.org/10.1007/BF02765184>
- Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology*, 74(6), 1421–1436. <https://doi.org/10.1037/0022-3514.74.6.1421>
- Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., & Graham, I. D. (2009). Evolution of Wenger's concept of community of practice. *Implementation Science*, 4, 11. <https://doi.org/10.1186/1748-5908-4-11>
- Likki, T., & Varazzani, C. (2017). *Applying behavioural insights to reduce pregnancy- and maternity-related discrimination view and disadvantage*. Equality and Human Rights Commission. <https://www.equalityhumanrights.com/en/publication-download/applying-behavioural-insights-reduce-pregnancy-and-maternity-related>
- Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, effectiveness, and ethics. *Basic and Applied Social Psychology*, 39(6), 293–306. <https://doi.org/10.1080/01973533.2017.1356304>
- Lindhout, P., & Reniers, G. (2017). What about nudges in the process industry? Exploring a new safety management tool. *Journal of Loss Prevention in the Process Industries*, 50, 243–256. <https://doi.org/10.1016/j.jlp.2017.10.006>
- Lodge, M., & Wegrich, K. (2016). The rationality paradox of nudge: Rational tools of government in a world of bounded rationality. *Law & Policy*, 38(3), 250–267. <https://doi.org/10.1111/lapo.12056>
- Loer, K. (2019). The enzymatic effect of behavioural sciences: What about policy-makers' expectations? In H. Straßheim, & S. Beck (Eds.), *Handbook of behavioural change and public policy* (pp. 375–405). Edward Elgar Publishing. <https://doi.org/10.4337/9781785367854.00020>
- Loo, R. (2002). A caveat on using single-item versus multiple-item scales. *Journal of Managerial Psychology*, 17(1), 68–75. <https://doi.org/10.1108/02683940210415933>
- Lorch, R. (1989). Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review*, 1(3), 209–234. <https://doi.org/10.1007/BF01320135>
- Lorch, R., Lorch, E., & Inman, W. E. (1993). Effects of signaling topic structure on text recall. *Journal of Educational Psychology*, 85(2), 281–290. <https://doi.org/10.1037/0022-0663.85.2.281>
- Loschelder, D. D., Siepelmeyer, H., Fischer, D., & Rubel, J. A. (2019). Dynamic norms drive sustainable consumption: Norm-based nudging helps café customers to avoid disposable to-go cups. *Journal of Economic Psychology*, 75, 102146. <https://doi.org/10.1016/j.joep.2019.02.002>
- Lourenço, J. S., Ciriolo, E., Almeida, S. R., & Troussard, X. (2016). *Behavioural insights applied to policy: European report 2016*. Publications Office of the European Union. <https://publications.jrc.ec.europa.eu/repository/handle/JRC100146>

- Ludwig, J., Kling, J. R., & Mullainathan, S. (2011). Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, 25(3), 17–38. <https://doi.org/10.1257/jep.25.3.17>
- Lunn, P. D. (2012). Behavioural economics and policymaking: Learning from the early adopters. *The Economic and Social Review*, 43(3), 423–449.
- Lupton, D. (2018). Towards design sociology. *Sociology Compass*, 12(1), e12546. <https://doi.org/10.1111/soc4.12546>
- Lynn, M. R. (2006). Sparks for sale: The culture and commerce of fireworks in early modern France. *Eighteenth-Century Life*, 30(2), 74–97. <https://doi.org/10.1215/00982601-2005-004>

## M

- Mahase, E. (2020). Covid-19: Was the decision to delay the UK's lockdown over fears of "behavioural fatigue" based on evidence? *BMJ*, 370, m3166. <https://doi.org/10.1136/bmj.m3166>
- Malpass, M. (2013). Between wit and reason: Defining associative, speculative, and critical design in practice. *Design and Culture*, 5(3), 333–356. <https://doi.org/10.2752/175470813X13705953612200>
- Manning, L., Dalton, A. G., Afif, Z., Vakos, R., & Naru, F. (2020). *Behavioral science around the world volume two: Profiles of 17 international organizations*. World Bank Group. <http://documents.worldbank.org/curated/en/453911601273837739/Behavioral-Science-Around-the-World-Volume-Two-Profiles-of-17-International-Organizations>
- Marchionni, C., & Reijula, S. (2019). What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science Part A*, 73, 54–63. <https://doi.org/10.1016/j.shpsa.2018.08.003>
- Marchiori, D. R., Adriaanse, M. A., & De Ridder, D. T. D. (2017). Unresolved questions in nudging research: Putting the psychology back in nudging. *Social and Personality Psychology Compass*, 11(1), e12297. <https://doi.org/10.1111/spc3.12297>
- Marshall, M., Pagel, C., French, C., Utley, M., Allwood, D., Fulop, N., Pope, C., Banks, V., & Goldmann, A. (2014). Moving improvement research closer to practice: The researcher-in-residence model. *BMJ Quality & Safety*, 23(10), 801–805. <https://doi.org/10.1136/bmjqs-2013-002779>
- Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: Can nudging improve population health? *BMJ*, 342, d228. <https://doi.org/10.1136/bmj.d228>
- Mattsson, J., Hedström, A., Viklander, M., & Blecken, G.-T. (2014). Fat, oil, and grease accumulation in sewer systems: Comprehensive survey of experiences of Scandinavian municipalities. *Journal of Environmental Engineering*, 140(3), 04014003. [https://doi.org/10.1061/\(asce\)ee.1943-7870.0000813](https://doi.org/10.1061/(asce)ee.1943-7870.0000813)
- McDavid, J., & Henderson, K. (2021). Hard won lessons learned in the evaluation field: Implications for behavioural insights practice. *Canadian Public Administration*, 64(1), 7–25. <https://doi.org/10.1111/capa.12401>



- McFarland, L. A., & Ployhart, R. E. (2015). Social media: A contextual framework to guide research and practice. *Journal of Applied Psychology*, 100(6), 1653–1677. <https://doi.org/10.1037/a0039244>
- McGann, M., loure, E., & Lewis, J. M. (2018). The rise of public sector innovation labs: Experiments in design thinking for policy. *Policy Sciences*, 51(3), 249–267. <https://doi.org/10.1007/s11077-018-9315-7>
- Mead, L. M. (2015). Only connect: Why government often ignores research. *Policy Sciences*, 48, 257–272. <https://doi.org/10.1007/s11077-015-9216-y>
- Meder, B., Fleischhut, N., & Osman, M. (2018). Beyond the confines of choice architecture: A critical analysis. *Journal of Economic Psychology*, 68, 36–44. <https://doi.org/10.1016/j.joep.2018.08.004>
- Mendel, J. (2018). Unpublished policy trials, the risk of false discoveries and the persistence of authority-based policy. *Evidence & Policy: A Journal of Research, Debate and Practice*, 14(2), 323–334. <https://doi.org/10.1332/174426416x14779397727921>
- Mewse, A. J., Lea, S. E. G., & Wrapson, W. (2010). First steps out of debt: Attitudes and social identity as predictors of contact by debtors with creditors. *Journal of Economic Psychology*, 31(6), 1021–1034. <https://doi.org/10.1016/j.joep.2010.08.009>
- Michie S., Atkins, L., & West, R. (2014). *The behaviour change wheel: A guide to developing interventions*. Silverback Publishing.
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6, 42. <https://doi.org/10.1186/1748-5908-6-42>
- Michie, S., West, R., & Harvey, N. (2020). The concept of “fatigue” in tackling covid-19. *BMJ*, 371, m4171. <https://doi.org/10.1136/bmj.m4171>
- Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, 13(3), 253–267. <https://doi.org/10.1080/09720502.2010.10700699>
- Miller, D. T., & Prentice, D. A. (2016). Changing norms to change behavior. *Annual Review of Psychology*, 67(1), 339–361. <https://doi.org/10.1146/annurev-psych-010814-01501>
- Mintzberg, H., & Waters, J. A. (1985). Of strategies, deliberate and emergent. *Strategic Management Journal*, 6(3), 257–72. <https://doi.org/10.1002/smj.4250060306>
- Mitchell, G. (2004). Libertarian paternalism is an oxymoron. *Northwestern University Law Review*, 99(3), 1245–1278.
- Mohapatra, B., & Mohamed, A. R. (2013). Place attachment and participation in management of neighbourhood green space: A place-based community management. *International Journal of Sustainable Society*, 5(3), 266–283. <https://doi.org/10.1504/ijssoc.2013.054715>
- Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, 20(8), 1045–1053. <https://doi.org/10.1177/1077800413513733>

- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48–76. <https://doi.org/10.1177/2345678906292462>
- Morris, M. (2004). Research on evaluation ethics: What have we learned and why is it important? *New Directions for Evaluation*, 1999(82), 15–24. <https://doi.org/10.1002/ev.1133>
- 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>
- Mukherjee, I., & Giest, S. (2020). Behavioural insights teams (BITs) and policy change: An exploration of impact, location, and temporality of policy advice. *Administration & Society*, 52(10), 1538–1561. <https://doi.org/10.1177/0095399720918315>
- Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Times Books/Henry Holt and Co.
- Mullin, M. (2021). Learning from local government research partnerships in a fragmented political setting. *Public Administration Review*, 81(5), 978–82. <https://doi.org/10.1111/puar.13395>
- Munafò, M. R., Nosek, B. A., Bishop, D., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021. <https://doi.org/10.1038/s41562-016-0021>
- Münscher, R., Vetter, M., & Scheuerle, T. (2016). A review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making*, 29(5), 511–524. <https://doi.org/10.1002/bdm.1897>
- Muramatsu, R., & Barbieri, F. (2017). Behavioral economics and Austrian economics: Lessons for policy and the prospects of nudges. *Journal of Behavioral Economics for Policy*, 1(1), 73–78.

## N

- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691–692. <https://doi.org/10.1093/biomet/78.3.691>
- Nagin, D. S., & Telep, C. W. (2017). Procedural justice and legal compliance. *Annual Review of Law and Social Science*, 13, 5–28. <https://doi.org/10.1146/annurev-lawsocsci-110316-113310>
- Nagtegaal, R., Tummers, L., Noordegraaf, M., & Bekkers, V. (2020). Designing to debias: Measuring and reducing public managers' anchoring bias. *Public Administration Review*, 80(4), 565–576. <https://doi.org/10.1111/puar.13211>
- Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behavior. *Safety Science*, 34(1–3), 99–109. [https://doi.org/10.1016/S0925-7535\(00\)00008-4](https://doi.org/10.1016/S0925-7535(00)00008-4)
- Neighborhood Profile Rotterdam (2016). *Wijkprofiel Rotterdam 2016*. <https://wijkprofiel.rotterdam.nl/>

- Niculescu, M., Payne, C. R., & Just, D. (2016). Using mirrors and social norms to improve fruit and vegetable purchases in grocery stores. *Proceedings of the 9th International Marketing Conference*, Romania, 9, 69–70.
- Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, 10, 53. <https://doi.org/10.1186/s13012-015-0242-0>
- Nocon, M., Müller-Riemenschneider, F., Nitzschke, K., & Willich, S. N. (2010). Review article: Increasing physical activity with point-of-choice prompts – A systematic review. *Scandinavian Journal of Public Health*, 38(6), 633–638. <https://doi.org/10.1177/1403494810375865>
- Nohlen H. U., & van Harreveld, F. (2017). *Psychologische inzichten in maatschappelijk (on)wenselijk gedrag: Sexting, voetbalvandalisme, winkeldiefstal, digitale geschiloplossing, en woonoverlast nader beschouwd*. Universiteit van Amsterdam.
- Nørgaard, A. S. (2018). Human behavior inside and outside bureaucracy: Lessons from psychology. *Journal of Behavioral Public Administration*, 1(1), 1–16. <https://doi.org/10.30636/jbpa.11.13>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., . . . Yarkoni, T. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Novemsky, N., Dhar, R., Schwarz, N., & Simonson, I. (2007). Preference fluency in choice. *Journal of Marketing Research*, 44(3), 347–356. <https://doi.org/10.1509/jmkr.44.3.347>
- Nunes, J. C., & Dreze, X. (2006). The endowed progress effect: How artificial advancement increases effort. *Journal of Consumer Research*, 32(4), 504–512. <https://doi.org/10.1086/500480>

## O

- OECD. (2017). *Behavioural insights and public policy: Lessons from around the world*. OECD Publishing. <https://doi.org/10.1787/9789264270480-en>
- OECD. (2019). *Tools and ethics for applied behavioural insights: The BASIC toolkit*. OECD Publishing. <https://doi.org/10.1787/9ea76a8f-en>
- Ogden, T., & Fixsen, D. L. (2014). Implementation science: A brief overview and a look ahead. *Zeitschrift für Psychologie*, 222(1), 4–11. <https://doi.org/10.1027/2151-2604/a000160>
- Olejniczak, K., Śliwowski, P., & Leeuw, F. (2020) Comparing behavioral assumptions of policy tools: Framework for policy designers. *Journal of Comparative Policy Analysis: Research and Practice*, 22(6), 498–520. <https://doi.org/10.1080/13876988.2020.1808465>

- Osman, M., McLachlan, S., Fenton, N., Neil, M., Löfstedt, R., & Meder, B. (2020). Learning from behavioural changes that fail. *Trends in Cognitive Sciences*, 24(12), 969–980. <https://doi.org/10.1016/j.tics.2020.09.009>
- Osman, M., Radford, S., Lin, Y., Gold, N., Nelson, W., & Löfstedt, R. (2018). Learning lessons: How to practice nudging around the world. *Journal of Risk Research*, 24(12), 969–980. <https://doi.org/10.1080/13669877.2018.1517127>
- Ovsiankina, M. (1928). Die Wiederaufnahme unterbrochener Handlungen. *Psychologische Forschung*, 11(1), 302–379. <https://doi.org/10.1007/BF00410261>

## P

- Park, W.-w. (1990). A review of research on groupthink. *Journal of Behavioral Decision Making*, 3(4), 229–245. <https://doi.org/10.1002/bdm.3960030402>
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Sage Publications.
- Pearce, W., & Raman, S. (2014). The new randomised controlled trials (RCT) movement in public policy: Challenges of epistemic governance. *Policy Sciences*, 47, 387–402. <https://doi.org/10.1007/s11077-014-9208-3>
- Pelz, D. C. (1978). Some expanded perspectives on use of social science in public policy. In Y. M. Yinger, & S. J. Cutler (Eds.), *Major social issues: A multidisciplinary view* (pp. 346–357). Free Press.
- Pirker, L., Gradišek, A., Višić, M., & Remškar, M. (2020). Nanoparticle exposure due to pyrotechnics during a football match. *Atmospheric Environment*, 233, 117567. <https://doi.org/10.1016/j.atmosenv.2020.117567>
- Plaks, J. E. (2017). Implicit theories: Assumptions that shape social and moral cognition. *Advances in Experimental Social Psychology*, 56, 259–310. <https://doi.org/10.1016/bs.aesp.2017.02.003>
- Politie (2018). *Jaarverslag voetbal en veiligheid: Seizoen 2017-2018*.
- Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2), 243–281. <https://doi.org/10.1111/j.1559-1816.2004.tb02547.x>

## R

- Reason, P. (Ed.). (1994). *Participation in human inquiry*. Sage Publications.
- Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy*, 6(1), 119–149. <http://doi.org/10.1017/bpp.2020.5>
- Rhodes, N., Shulman, H. C., & McClaran, N. (2020). Changing norms: A meta-analytic integration of research on social norms appeals. *Human Communication Research*, 46(2–3), 161–191. <https://doi.org/10.1093/hcr/hqz023>
- Rich, R. F. (1997). Measuring knowledge utilization: Processes and outcomes. *Knowledge and Policy*, 10, 11–24. <https://doi.org/10.1007/bf02912504>

- Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2007). Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology*, 43(6), 933–946.  
<https://doi.org/10.1016/j.jesp.2006.10.014>
- Roach, J., Cartwright, A., Weir, K., Richards, S., & Weir, M. (2020). Reducing student burglary victimization using the nudge approach. *Crime Prevention and Community Safety*, 22(4), 364–380. <https://doi.org/10.1057/s41300-020-00100-y>
- Roda, C., Charreire, H., Feuillet, T., Mackenbach, J. D., Compernelle, S., Glonti, K., Ben Rebah, M., Bárdos, H., Rutter, H., McKee, M., De Bourdeaudhuij, I., Brug, J., Lakerveld, J., & Oppert, J.-M. (2016). Mismatch between perceived and objectively measured environmental obesogenic features in European neighbourhoods. *Obesity Reviews*, 17(S1), 31–41.  
<https://doi.org/10.1111/obr.12376>
- Rogers, T., & Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour*, 2(5), 335–342. <https://doi.org/10.1038/s41562-018-0328-1>
- Rogers, E., & Weber, E. P. (2010). Thinking harder about outcomes for collaborative governance arrangements. *The American Review of Public Administration*, 40(5), 546–567.  
<https://doi.org/10.1177/0275074009359024>
- Rookwood, J., & Pearson, G. (2012). The hoolifan: Positive fan attitudes to football 'hooliganism'. *International Review for the Sociology of Sport*, 47(2), 149–164.  
<https://doi.org/10.1177/1012690210388455>
- Rosseel Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Ruberton, P. M., Gladstone, J., & Lyubomirsky, S. (2016). How your bank balance buys happiness: The importance of “cash on hand” to life satisfaction. *Emotion*, 16(5), 575–580. <https://doi.org/10.1037/emo0000184>

## S

- Sabatier, P. A., & Weible, C. M. (2007). The advocacy coalition framework. In P. A. Sabatier (Ed.), *Theories of the policy process* (pp. 189–222). Routledge.  
<https://doi.org/10.4324/9780367274689>
- Salamon, L. M. (2002). *The tools of government: A guide to the new governance*. Oxford University Press.
- 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>
- Saunders, R. P., Evans, M. H., & Joshi, P. (2005). Developing a process-evaluation plan for assessing health promotion program implementation: A how-to guide. *Health Promotion Practice*, 6(2), 134–147. <https://doi.org/10.1177/1524839904273387>

- Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, 54(3), 182–203. <https://doi.org/10.1037/0003-066X.54.3.182>
- Schmidt, R., & Stenger, K. (forthcoming). Behavioral brittleness: The case for strategic behavioral public policy. *Behavioural Public Policy*. <https://doi.org/10.1017/bpp.2021.16>
- Schneider, A., & Ingram, H. (1990). Behavioral assumptions of policy tools. *The Journal of Politics*, 52(2), 510–529. <https://doi.org/10.2307/2131904>
- Scholz P. (2020). Female visitors and their behavior at football stadiums: Czechia case study. *Journal of Physical Education and Sport*, 20(4), 1803–1808. <https://doi.org/10.7752/jpes.2020.04244>
- Scholz, P., & Hurych, E. (2018). The ethical issue in the case of selected football supporters: Deontology or consequentialism? *Journal of Physical Education and Sport*, 18(3), 1678–1684. <https://doi.org/10.7752/jpes.2018.03245>
- Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.
- Schultz, P., Bator, R. J., Large, L. B., Bruni, C. M., & Tabanico, J. J. (2013). Littering in context: Personal and environmental predictors of littering behavior. *Environment and Behavior*, 45(1), 35–59. <https://doi.org/10.1177/0013916511412179>
- Schwartz, S. H. (1977). Normative influences on altruism. *Advances in Experimental Social Psychology*, 10, 221–279. [https://doi.org/10.1016/S0065-2601\(08\)60358-5](https://doi.org/10.1016/S0065-2601(08)60358-5)
- Selinger, E., & Whyte, K. P. (2012). Nudging cannot solve complex policy issues. *European Journal of Risk Regulation*, 3(1), 26–31. <https://doi.org/10.1017/s1867299x0000177x>
- Service, O., Hallsworth, M., Halpern, D., Algate, F., Gallagher, R., Nguyen, S., Ruda, S., Sanders, M., Pelenur, M., Gyani, A., Harper, H., Reinhard, J., & Kirkman, E. (2014). *EAST: Four simple ways to apply behavioural insights*. Cabinet Office Behavioural Insights Team.
- Shah, A. K., Mullainathan S., & Shafir, E. (2012). Some consequences of having too little. *Science*, 338(6107), 682–85. <https://doi.org/10.1126/science.1222426>
- Shah, A. K., Shafir, E., & Mullainathan, S. (2015). Scarcity frames value. *Psychological Science*, 26(4), 402–412. <https://doi.org/10.1177/0956797614563958>
- Shah, A. K., Shafir, E., & Mullainathan, S. (2019). An exercise in self-replication: Replicating Shah, Mullainathan, and Shafir (2012). *Journal of Economic Psychology*, 75, 102127. <https://doi.org/10.1016/j.joep.2018.12.001>
- Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. *Social and Personality Psychology Compass*, 10(9), 503–518. <https://doi.org/10.1111/spc3.12265>
- Sheridan, J. E. (1992). Organizational culture and employee retention. *The Academy of Management Journal*, 35(5), 1036–1056. <https://doi.org/10.2307/256539>
- Simon, H. A. (1957). *Models of man: Social and rational*. Wiley.
- Smith, K. E., Bandola-Gill, J., Meer, N., Stewart E., & Watermeyer R. (2020). *The impact agenda: Controversies, consequences and challenges*. Policy Press.

- Smith, S., Ward, V., & House, Al. (2011). 'Impact' in the proposals for the UK's Research Excellence Framework: Shifting the boundaries of academic autonomy. *Research Policy*, 40(10), 1369–1379. <https://doi.org/10.1016/j.respol.2011.05.026>
- Sofaer, S. (1999), Qualitative methods: What are they and why use them? *Health Services Research*, 34(5), 1101–1118.
- Soler, R. E., Leeks, K. D., Buchanan, L. R., Brownson, R. C., Heath, G. W., & Hopkins, D. H. (2010). Point-of-decision prompts to increase stair use: A systematic review update. *American Journal of Preventive Medicine*, 38(Suppl. 2), S292–S300. <https://doi.org/10.1016/j.amepre.2009.10.028>
- Soman, D., & Yeung, C. (Eds.). (2021). *The behaviorally informed organization*. University of Toronto Press.
- Sommestad, T., Hallberg, J., Lundholm, K., & Bengtsson, J. (2014). Variables influencing information security policy compliance: A systematic review of quantitative studies. *Information Management & Computer Security*, 22(1), 42–75. <https://doi.org/10.1108/IMCS-08-2012-0045>
- Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science*, 19(10), 986–988. <https://doi.org/10.1111/j.1467-9280.2008.02189.x>
- Spaaij, R. (2008). Men like us, boys like them: Violence, masculinity, and collective identity in football hooliganism. *Journal of Sport and Social Issues*, 32(4), 369–392. <https://doi.org/10.1177/0193723508324082>
- Sparkman, G., & Walton, G. M. (2017). Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological Science*, 28(11), 1663–1674. <https://doi.org/10.1177/0956797617719950>
- Stake, R.E., & Trumbull, D. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science*, 7(1), 1-12.
- Standing Committee of the European Convention on Spectator Violence. (2011). *European convention on spectator violence and misbehaviour at sports events and in particular football matches*. Council of Europe.
- Steckler, A. B., & Linnan, L. (2002). *Process evaluation for public health interventions and research*. Jossey-Bass.
- Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology*, 49(3), 607–627. <https://doi.org/10.1037/0022-3514.49.3.607>
- Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology*, 41(1), 37–55. <https://doi.org/10.1037/0022-3514.41.1.37>
- Stott, C., & Adang, O. (2009). *Policing football matches with an international dimension in the European Union: Understanding and managing risk*. Bavnbank Press.

- Strassheim, H. (2019). Behavioural mechanisms and public policy design: Preventing failures in behavioural public policy. *Public Policy and Administration*, 36(2), 187–204.  
<https://doi.org/10.1177/0952076719827062>
- Strassheim, H. (2020a). De-biasing democracy: Behavioural public policy and the post-democratic turn. *Democratization*, 27(3), 461–476. <https://doi.org/10.1080/13510347.2019.1663501>
- Strassheim, H. (2020b). The rise and spread of behavioral public policy: An opportunity for critical research and self-reflection. *International Review of Public Policy*, 2(1), 115–128.  
<https://doi.org/10.4000/irpp.897>
- Strassheim, H. (2021). Who are behavioural public policy experts and how are they organised globally? *Policy & Politics*, 49(1), 69–86. <https://doi.org/10.1332/030557320x15956825120821>
- Strassheim, H., & Beck, S. (Eds.). (2019). *Handbook of behavioural change and public policy*. Edward Elgar Publishing. <https://doi.org/10.4337/9781785367854>
- Strassheim, H., Jung, A., & Korinek, R.-L. (2015). Reframing expertise: The rise of behavioural insights and interventions in public policy. In A. Berthoin Antal, M. Hutter, & D. Stark (Eds.), *Moments of valuation: Exploring sites of dissonance* (pp. 249–268). Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780198702504.003.0013>
- Stroebe, W., Nijstad, B. A., & Rietzschel, E. F. (2010). Beyond productivity loss in brainstorming groups: The evolution of a question. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 43, pp. 157–203). Academic Press. [https://doi.org/10.1016/S0065-2601\(10\)43004-X](https://doi.org/10.1016/S0065-2601(10)43004-X)
- Sunstein, C. R. (2014). Nudging: A very short guide. *Journal of Consumer Policy*, 37(4), 583–588.  
<https://doi.org/10.1007/s10603-014-9273-1>
- Sunstein, C. R. (2016). *The ethics of influence: Government in the age of behavioral science*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316493021>
- Sunstein, C. R. (2017). Nudges that fail. *Behavioural Public Policy*, 1(1), 4–25.  
<https://doi.org/10.1017/bpp.2016.3>
- Sunstein, C. R. (forthcoming). Sludge audits. *Behavioural Public Policy*.  
<https://doi.org/10.1017/bpp.2019.32>
- Sutherland, R. (2018). Policy-making under uncertainty. *Behavioural Public Policy*, 2(2), 246–251.  
<https://doi.org/10.1017/bpp.2018.21>
- Sweet, E., Nandi, A., Adam, E. K., & McDade, T. W. (2013). The high price of debt: Household financial debt and its impact on mental and physical health. *Social Science & Medicine*, 91, 94–100. <https://doi.org/10.1016/j.socscimed.2013.05.009>
- Swim, J. K., Geiger, N., & Zawadzki, S. J. (2014). Psychology and energy-use reduction policies. *Policy Insights from the Behavioral and Brain Sciences*, 1(1), 180–188.  
<https://doi.org/10.1177/2372732214548591>
- Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, 31(3), 355–366. <https://doi.org/10.1002/bdm.2035>



## T

- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin (Ed.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.
- Tay, L., Batz, C., Parrigon, S., & Kuykendall, L. (2017). Debt and subjective well-being: The other side of the income-happiness coin. *Journal of Happiness Studies*, 18(3), 903–937.  
<https://doi.org/10.1007/s10902-016-9758-5>
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, 93(2), 175–179. <https://doi.org/10.1257/000282803321947001>
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- Thaler, R. H., & Sunstein, C. R. (2021). *Nudge: The final edition*. Penguin.
- Thomas, M. D. (2019). Reapplying behavioral symmetry: Public choice and choice architecture. *Public Choice*, 180, 11–25. <https://doi.org/10.1007/s11127-018-0537-1>
- Tiemeijer, W. (2016). *Eigen schuld? Een gedragswetenschappelijk perspectief op problematische schulden*. Amsterdam University Press.
- Tor, A. (2020). Nudges that should fail? *Behavioural Public Policy*, 4(3), 316–342.  
<https://doi.org/10.1017/bpp.2019.5>
- Torring, J. (2019). Collaborative innovation in the public sector: The argument. *Public Management Review*, 21(1), 1–11. <https://doi.org/10.1080/14719037.2018.1430248>
- Tuckman, B. W., & Jensen, M. A. C. (1977). Stages of small-group development revisited. *Group & Organization Studies*, 2(4), 419–427. <https://doi.org/10.1177/105960117700200404>
- Tummers, L., & Bekkers, V. (2014). Policy implementation, street-level bureaucracy, and the importance of discretion. *Public Management Review*, 16(4), 527–47.  
<https://doi.org/10.1080/14719037.2013.841978>
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

## U

- UN Innovation Network. (2021). *United Nations behavioral science report*.  
[https://www.uninnovation.network/assets/BeSci/UN\\_Behavioural\\_Science\\_Report\\_2021.pdf](https://www.uninnovation.network/assets/BeSci/UN_Behavioural_Science_Report_2021.pdf)

## V

- van Buuren, A., Lewis, J. M., Guy Peters, B., & Voorberg, W. (2020). Improving public policy and administration: Exploring the potential of design. *Policy & Politics*, 48(1), 3–19.  
<https://doi.org/10.1332/030557319X15579230420063>
- van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, 16(2), 75–84.

- Van De Ven, A. H., & Johnson, P. E. (2006). Knowledge for theory and practice. *The Academy of Management Review*, 31(4), 802–821. <https://doi.org/10.2307/20159252>
- van Dierendonck D. (2018) Building a people-oriented and servant-led organization. In D. van Dierendonck, & K. Patterson (Eds.) *Practicing servant leadership* (pp. 169–186). Palgrave Macmillan. [https://doi.org/10.1007/978-3-319-75644-8\\_10](https://doi.org/10.1007/978-3-319-75644-8_10)
- van Ham, T., Adang, O. M. J., Ferwerda, H. B., Doreleijers, T. A. H., & Blokland, A. A. J. (forthcoming). Planned hooligan fights: Contributing factors and significance for individuals who take part. *European Journal of Criminology*. <https://doi.org/10.1177/1477370820932080>
- van Kleef, E., Seijdel, K., Vingerhoeds, M. H., de Wijk, R. A., & van Trijp, H. C. M. (2018). The effect of a default-based nudge on the choice of whole-wheat bread. *Appetite*, 121, 179–185. <https://doi.org/10.1016/j.appet.2017.11.091>
- van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review*, 25(1), 143–154. <https://doi.org/10.3758/s13423-016-1015-8>
- Van Ryzin, G. G. (2021). Nudging and muddling through. *Perspectives on Public Management and Governance*, 4(4), 339–345. <https://doi.org/10.1093/ppmgov/gvab021>
- Van Vugt, M. (2009). Averting the tragedy of the commons: Using social psychological science to protect the environment. *Current Directions in Psychological Science*, 18(3), 169–173. <https://doi.org/10.1111/j.1467-8721.2009.01630.x>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16(2), 667–718. <https://doi.org/10.1214/20-BA1221>
- Verbeeten, F. H. M., & Spekklé, R. F. (2015). Management control, results-oriented culture and public sector performance: Empirical evidence on new public management. *Organization Studies*, 36(7), 953–978. <https://doi.org/10.1177/0170840615580014>
- Vis, B. (2019). Heuristics and political elites' judgment and decision-making. *Political Studies Review*, 17(1), 41–52. <https://doi.org/10.1177/1478929917750311>
- Vivian, J. E., & Berkowitz, N. H. (1992). Anticipated bias from an outgroup: An attributional analysis. *European Journal of Social Psychology*, 22(4), 415–424. <https://doi.org/10.1002/ejsp.2420220410>
- Voß, J.-P., & Simons, A. (2018). A Novel understanding of experimentation in governance: Co-producing innovations between “lab” and “field”. *Policy Sciences*, 51, 213–229. <https://doi.org/10.1007/s11077-018-9313-9>

## W

- Walker, C. (2012). Personal debt, cognitive delinquency and techniques of governmentality: Neoliberal constructions of financial inadequacy in the UK. *Journal of Community & Applied Social Psychology*, 22(6), 533–538. <https://doi.org/10.1002/casp.1127>

- Walker, C., Burton, M., Akhurst, J., & Degirmencioglu, S. M. (2015). Locked into the system? Critical community psychology approaches to personal debt in the context of crises of capital accumulation. *Journal of Community & Applied Social Psychology*, 25(3), 264–275. <https://doi.org/10.1002/casp.2209>
- Walter, I., Davies, H., & Nutley, S. (2003). Increasing research impact through partnerships: Evidence from outside health care. *Journal of Health Services Research & Policy*, 8(2S), 58–61. <https://doi.org/10.1258/135581903322405180>
- Wang, C. S., Kenneth, T., Ku, G., & Galinsky, A. D. (2014). Perspective-taking increases willingness to engage in intergroup contact. *PLoS ONE*, 9(1), e85681. <https://doi.org/10.1371/journal.pone.0085681>
- Ward, R. E., Jr. (2002). Fan violence: Social problem or moral panic? *Aggression and Violent Behavior*, 7(5), 453–475. [https://doi.org/10.1016/S1359-1789\(01\)00075-1](https://doi.org/10.1016/S1359-1789(01)00075-1)
- Ward, V. (2017). Why, whose, what and how? A framework for knowledge mobilisers. *Evidence & Policy: A Journal of Research, Debate and Practice*, 13(3), 477–497. <https://doi.org/10.1332/174426416X14634763278725>
- Ward, V., Smith, S., House, A., & Hamer, S. (2012). Exploring knowledge exchange: a useful framework for practice and policy. *Social Science & Medicine*, 74(3), 297–304. <https://doi.org/10.1016/j.socscimed.2011.09.021>
- Ward, V., Tooman, T., Reid, B., Davies, H., & Marshall, M., (2021a). Embedding researchers into organisations: A study of the features of embedded research initiatives. *Evidence & Policy: A Journal of Research, Debate and Practice*, 17(4), 593–614. <https://doi.org/10.1332/174426421X16165177580453>
- Ward, V., Tooman, T., Reid, B., Davies, H., O'Brien, B., Mear, L., & Marshall, M. (2021b). A framework to support the design and cultivation of embedded research initiatives. *Evidence & Policy: A Journal of Research, Debate and Practice*, 17(4), 755-769. <https://doi.org/10.1332/174426421X16165177707227>
- Warkentin, M., & Willison, R. (2009). Behavioral and policy issues in information systems security: The insider threat. *European Journal of Information Systems*, 18(2), 101–105. <https://doi.org/10.1057/ejis.2009.12>
- Weaver, R. K. (2014). Compliance regimes and behavioral change. *Governance*, 27(2), 243–265. <https://doi.org/10.1111/gove.12032>
- Weaver, R. K. (2015). Getting people to behave: Research lessons for policy makers. *Public Administration Review*, 75(6), 806–816. <https://doi.org/10.1111/puar.12412>
- Weible, C. M., Heikkilä, T., deLeon, P., & Sabatier, P. A. (2012). Understanding and influencing the policy process. *Policy Sciences*, 45(1), 1–21. <https://doi.org/10.1007/s11077-011-9143-5>
- Weijers, R. J., & de Koning, B. B. (2021). Nudging to increase hand hygiene during the COVID-19 pandemic: A field experiment. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 53(3), 353-357. <http://dx.doi.org/10.1037/cbs0000245>

- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803932>
- Wetenschappelijke Raad voor het Regeringsbeleid. (2014). *Met kennis van gedrag beleid maken* (Report No. 92). Amsterdam University Press.
- Whitehead, M., Jones, R., Lilley, R., Howell, R., & Pykett, J. (2018). Neuroliberalism: Cognition, context, and the geographical bounding of rationality. *Progress in Human Geography*, 43(4), 632–649. <https://doi.org/10.1177/0309132518777624>
- Whitehead, M., Jones, R., Lilley, R., Pykett, J., & Howell, R. (2017). *Neuroliberalism: Behavioural government in the twenty-first century*. Routledge. <https://doi.org/10.4324/9781315684772>
- Wicklund, R. A. (1979). The influence of self-awareness on human behavior: The person who becomes self-aware is more likely to act consistently, be faithful to societal norms, and give accurate reports about himself. *American Scientist*, 67(2), 187–193.
- World Bank. (2015). *World development report 2015: Mind, society, and behavior*. World Bank Group. <https://doi.org/10.1596/978-1-4648-0342-0>
- Wu, S. J., & Paluck, E. L. (2021). Designing nudges for the context: Golden coin decals nudge workplace behavior in China. *Organizational Behavior and Human Decision Processes*, 163, 43–50. <https://doi.org/10.1016/j.obhdp.2018.10.002>

## Y

- Yazdanmehr, A., & Wang, J. (2016). Employees' information security policy compliance: A norm activation perspective. *Decision Support Systems*, 92, 36–46. <https://doi.org/10.1016/j.dss.2016.09.009>

## Z

- Zhang, C. Y., Hemmeter, J., Kessler, J. B., Metcalfe, R. D., & Weathers, R. (2020). *Nudging timely wage reporting: Field experimental evidence from the United States Social Supplementary Income Program* (Working Paper No. 27875). National Bureau of Economic Research. <https://doi.org/10.3386/w27875>
- Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206. <https://doi.org/10.1086/651257>
- Zhao, J., & Tomm, B. M. (2017). Attentional trade-offs under resource scarcity. In D. D. Schmorow & C. M. Fidopiastis (Eds.). *Augmented cognition: Enhancing cognition and behavior in complex human environments* (pp. 78–97). Springer. [https://doi.org/10.1007/978-3-319-58625-0\\_6](https://doi.org/10.1007/978-3-319-58625-0_6)
- Zimniak, R. (2020). The sociological and psychological aspect of football hooliganism. *Teisé*. 2020, 117, 138–151. <https://doi.org/10.15388/Teise.2020.117.10>
- Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology*, 65(1), 96–102. <https://doi.org/10.1037/0021-9010.65.1.96>

## Curriculum Vitae

Malte Dewies (1991) grew up with his family in the rural area of Mönchengladbach, Germany. As a child, he loved playing with Lego bricks and with the neighbourhood kids outside. Since being a teenager, Malte enjoyed acting and performed on stage regularly. What he enjoyed most about acting was to investigate the human mind and why humans act the way they do. Ultimately, this motivated him to study psychology. He started his studies at Ruhr-University Bochum, Germany, in 2012. Being interested in a career in business consulting, he chose to specialise in organisational psychology.

During his bachelor studies, Malte was able to spend three months at the University of Birmingham, England. In this new environment, he experienced different methods of teaching and got fresh insights that he could not have gotten when staying at Ruhr-University Bochum. Seeking more novel experiences, Malte was convinced that switching to a different university for his master was a great idea, and from 2015 he continued his studies at the University of Mannheim, Germany. After one year, he went to study abroad at the KU Leuven, Belgium. By then, Malte had concluded that working in the commercial sector could not sufficiently satisfy his thirst for knowledge and his felt need to contribute to society. Therefore, he applied at universities to become a PhD student.

Malte started his PhD at Erasmus University Rotterdam in February 2018. For this, he collaborated with the municipality of Rotterdam to investigate how a better understanding of human behaviour could be integrated into public policy. In 2021, he was awarded with a full PhD scholarship from the German Academic Scholarship Foundation.

# Portfolio

## Publications

Scientific papers published or accepted for publication in peer-reviewed academic journals.

**Dewies, M.**, Merkelbach, I., van der Scheer, W. K., Rohde, K. I. M., & Denktas, S. (forthcoming). Do you consider human behaviour to be stable or malleable? Your answer can influence your preferences for policy instruments. *Journal of Behavioral Public Administration*.

**Dewies, M.**, Schop-Etman, A., Merkelbach, I., Rohde, K. I. M., & Denktas, S. (forthcoming). Call first, pay later: Stimulating debtors to contact their creditors improves debt collection in the context of financial scarcity. *Behavioural Public Policy*.

**Dewies, M.**, Denktas, S., Giel, L., Noordzij, G., & Merkelbach, I. (2022). Applying behavioural insights to public policy: An example from Rotterdam. *Global Implementation Research and Applications*, 2(1), 53–66. <https://doi.org/10.1007/s43477-022-00036-5>

**Dewies, M.**, Schop-Etman, A., Rohde, K. I. M., & Denktas, S. (2021). Nudging is ineffective when attitudes are unsupportive: An example from a natural field experiment. *Basic and Applied Social Psychology*, 43(4), 213-225. <https://doi.org/10.1080/01973533.2021.1917412>

Merkelbach, I., **Dewies, M.**, & Denktas, S. (2021). Committing to keep clean: Nudging complements standard policy measures to reduce illegal urban garbage disposal in a neighborhood with high levels of social cohesion. *Frontiers in Psychology*, 12, 660410. <https://doi.org/10.3389/fpsyg.2021.660410>

## Scientific Papers

Scientific papers in progress.

Merkelbach, I., **Dewies, M.**, Noordzij, G., & Denktas, S. (2021). No pyro, no party: Social factors, deliberate choices, and shared fan culture determine the use of illegal fireworks in a soccer stadium [version 1; peer review: awaiting peer review]. *F1000Research*.

**Dewies, M.**, Merkelbach, I., Edelenbos, J., Rohde, K.I.M., & Denktas, S. (submitted). Comprehensive evaluation of the Behavioural Insights Group Rotterdam.

**Dewies, M.**, Merkelbach, I., Rohde, K.I.M., & Denktas, S. (submitted). Strategies for behavioural insights teams.

## Professional Publication

Publication in a professional journal.

Merkelbach, I., van den Boogaard, L., Zegveld-Steur, M., **Dewies, M.**, & Kooijman, M. (2020). Gedrag sturen in crisistijd. *Binnenlands Bestuur*, 8, 22-24.



## Award

Full PhD scholarship from the German Academic Scholarship Foundation (for more information please visit <https://www.studienstiftung.de/en/doctoral-scholarships/>).

## Presentations

Presentations for professionals and researchers at (inter)national conferences and symposia.

Presenting author(s) are indicated with an asterisk (“\*”).

**Dewies, M.\***, Merkelbach, I., & Denktas, S. (2022, January). *Approach and evaluation of the Behavioural Insights Group Rotterdam*. Online presentation at the “ROCKS meeting” of the “Erasmus Clinical & Health Psychology” group at Erasmus University Rotterdam, the Netherlands.

**Dewies, M.\***, Merkelbach, I., Edelenbos, J., Rohde, K. I. M., & Denktas, S. (2021, November). *Evaluation of the Behavioural Insights Group Rotterdam*. Presentation at the annual conference of the “Netherlands Institute of Governance” in Utrecht, the Netherlands.

**Dewies, M.\***, Merkelbach, I., Rohde, K. I. M., van der Scheer, W., & Denktas, S. (2021, October). *Beleid en gedrag [Policy and behaviour]*. Online presentation at the “Vereniging Gehandicaptenzorg Nederland” [Dutch association of healthcare providers for people with disabilities], the Netherlands.

**Dewies, M.\***, Schop-Etman, A., Merkelbach, I., Rohde, K. I. M., & Denktas, S. (2021, June). *Call first, pay later: Nudging debt repayment*. Online presentation at the annual conference of the “International Association for Research in Economic Psychology (IAREP)” and the “Society for the Advancement of Behavioral Economics (SABE)”.

**Dewies, M.\***, Schop-Etman, A., Rohde, K. I. M., Merkelbach, I., & Denktas, S. (2021, June). *Ineffective nudging*. Online presentation at the „DPECS Graduate Research Day 2021” at Erasmus University Rotterdam, the Netherlands.

**Dewies, M.\***, Merkelbach, I., Rohde, K. I. M., & Denktas, S. (2021, May). *Can people change their behaviour? You answer can influence your policy choices*. Online presentation at the “Resilience and recovery after Covid-19” conference of the “Erasmus School of Social and Behavioural Sciences” at Erasmus University Rotterdam, the Netherlands.

**Dewies, M.\*** (2020, October). *Aanpak en evaluatie van de Behavioural Insights Group Rotterdam [Approach and evaluation of the Behavioural Insights Group Rotterdam]*. Online presentation at the “Gedragstnetwerk Werk en Inkomen” [Behaviour insights network Work and Income] for policy professionals at the Municipality of Rotterdam, the Netherlands.

**Dewies, M.\***, & Merkelbach, I. & Denktas, S. (2020, September). *Gedragsteams bij de overheid: Sturen op gedrag [Behavioural Insights Teams at governments: Influencing behaviour]*. Presentation at the “Gedragbestuurskunde” [Behavioural public administration] symposium of

the “BOW Studievereniging” [BOW student association] at the University Nijmegen, the Netherlands.

**Dewies, M.\***, & Schop-Etman, A. (2020, February). *Food waste recycling: A BIG'R case study*. Presentation at the “Co-creation session: Waste separation in high rise student flats” of ErasmusX at the Erasmus University Rotterdam, the Netherlands.

**Dewies, M.\***, Schop-Etman, A., & Denktas, S. (2020, February). *Getting in touch with poor debtors*. Online presentation at the “ROCKS meeting” of the “Erasmus Clinical & Health Psychology” group at Erasmus University Rotterdam, the Netherlands.

**Dewies, M.\***, Janssen, A\*, & Stam, B. (2020, January). *De BIG'R concernpas casus: resultaten en implicaties [The BIG'R employee badge case: Results and implications]*. Presentation at the “Brown bag lunch” of the Safety Department at the Municipality of Rotterdam, the Netherlands.

**Dewies, M.\***, & Schäfer, A. (2019, November). *De BIG'R casus: vet in het riool [The BIG'R case: Fat in the sewer]*. Presentation at the “Gedragskijkers” [Behaviour-viewers] of the City Maintenance subdivision at the Municipality of Rotterdam, the Netherlands.

**Dewies, M.\***, Robbers, S., Rohde, K. I. M., & Denktas, S. (2019, June). *Employing EAST for debt repayment letters: Lessons learned from the Behavioural Insights Group Rotterdam (BIGR) initiated field experiment*. Presentation at the “Nudging and Beyond” conference of the “Welfare Improvement through Nudging Knowledge (WINK)” consortium at Utrecht University, the Netherlands.

**Dewies, M.\***, Schop-Etman, A., & Denktas, S. (2019, November). *Aanpak en resultaten van de Behavioural Insights Group Rotterdam [Approach and results from the Behavioural Insights Group Rotterdam]*. Presentation at the “Dag van het gedrag” [Day of behaviour] annual conference of the “Behavioural Insights Network Netherlands” in The Hague, the Netherlands.

Woodend, A., Giel, L., Robbers, S.\*, **Dewies, M.**, Kooijman, M., Watzeels, A. J. C. M., Wolfers, M. E. G., von Harras, M., Noordzij, G., & Denktas, S. (2018, October). *Behavioural Insights Group Rotterdam*. Poster presentation at the “New Directions in Psychology of Behavior Change” symposium of the “Behavioral Science Institute” at Radboud University, the Netherlands.

Woodend, A.\*, Giel, L., Robbers, S., **Dewies, M.\***, Kooijman, M., Watzeels, A. J. C. M., Wolfers, M. E. G., von Harras, M., Noordzij, G., & Denktas, S. (2018, September). *Behavioural Insights Group Rotterdam*. Poster presentation at the “Science meets city” knowledge festival of the Erasmus University Rotterdam, the Netherlands.

Goijvaerts, E.\*, **Dewies, M.\***, & Robbers, S. (2018, June). *Aanpak van de Behavioural Insights Group Rotterdam [Approach of the Behavioural Insights Group Rotterdam]*. Presentation at the “Impact door samenwerking [Impact by collaboration]” annual congress of the “Vereniging voor Statistiek en Onderzoek” [Association for statistics and research] in Amersfoort, the Netherlands.

## Workshops and Trainings

Workshop and trainings to improve the capabilities of professionals.

**Dewies, M., & Merkelbach, I.** (2021, December). *Implementation of behavioural interventions*. Lecture for the “Nudging: sturen op gedrag” [Nudging: Changing behaviours] professional course at Erasmus Academy at Erasmus University in Rotterdam, the Netherlands.

Kooijman, M., Slag, G., **Dewies, M.**, & Mamede A., (2020, August). *BIG'R zomerschool: met kennis van gedrag beleid maken* [BIG'R summer school: Policy making using knowledge about behaviour]. Online training session at the municipality of Rotterdam, the Netherlands.

Kooijman, M., Slag, G., **Dewies, M.**, & Schäfer, A. (2020, July). *BIG'R zomerschool: met kennis van gedrag beleid maken* [BIG'R summer school: Policy making using knowledge about behaviour]. Online training session at the municipality of Rotterdam, the Netherlands.

Merkelbach, I., & **Dewies, M.** (2020, April). *Verander menselijk gedrag tijdens de CORONA crisis* [Changing human behaviour during the Covid-19 crisis]. Webinar for the Erasmus Academy at Erasmus University in Rotterdam, the Netherlands

Denktaş, S., Robbers, S., **Dewies, M.**, Noordzij, G., & Schop-Etman, A. (2019, January). *Sturen met psychologie* [Governing with psychology]. Workshop at the “Sturen op transitie: hoe doe je dat?” [Governing transitions: How to?] knowledge fair at the Erasmus University Rotterdam, the Netherlands.

Denktaş, S., Goijvaerts, E., **Dewies, M.**, van Rijn, P., & Kooijman, M. (2018, December). *Omgeving en menselijk gedrag* [Context and human behaviour]. Workshop at a local business summit from the municipality of Rotterdam in Rotterdam, the Netherlands.

Robbers, S., **Dewies, M.**, van den Boogaard, L. Goijvaerts, E., & van den Berge, S. (2018, July). *Met gedragsinzichten de opkomst van jonge stemmers verhogen* [Using behavioural insights to increase young voters' turnout]. Workshop at the “Samen Noord-Holland: Toekomst Congres” [Together North-Holland: Future conference] for the Government of the Province North Holland in Haarlem, the Netherlands.

## Policy Advices

Policy advices produced for the Behavioural Insights Group Rotterdam.

Christiaanse, B., Slag, G., & **Dewies, M.** (2020). *De “WW@WORK” casus* [The “WW@WORK” case].

**Dewies, M.**, Christiaanse, B., Schop-Etman, A., Vogels, M., & Zegveld-Steur, M. (2019). *The “Conradstraat” case* [The “Conradstraat” case].

**Dewies, M.**, Merkelbach, I., Christiaanse, B., van Rijn, P., & Zegveld-Steur, M. (2020). *Brief Schuldenbeleid - Kan een met nudges verrijkte brief contact bevorderen tussen Bureau Frontlijn en Rotterdammers met een betalingsachterstand op de zorgpremie?* [Indebtedness letter: Can a

*letter that includes nudges stimulate contact between Bureau Frontlijn and Rotterdam residents with delayed health insurance payments].*

**Dewies, M.,** Merkelbach, I., & Kooijman, M. (2020). *Afspraakbevestiging voor een corona-proof huisbezoek van Bouw- en Woningtoezicht [Confirmation notice for a Covid-19-proof house visit from Construction and Housing Control].*

**Dewies, M.,** Stam, B., & Kooijman, M. (2020). *De “laadpaalkleven” case [The “charging station hogging” case].*

**Dewies, M.,** Stam, B., Schop-Etman, A., Vogels, M., & Zegveld-Steur, M. (2019). *De “vet in het riool” casus [The “fat in the sewer” case].*

Kooijman, M., **Dewies, M.,** Robbers, S., Middelkoop, R., Wolfers, M. E. G., Giel, L., Noordzij, G., & Denktas, S. (2018). *De OV-knooppunt Zuidplein casus [The Zuidplein public transport hub case].*

Kooijman, M., Robbers, S., & **Dewies, M.** (2018). *De “maatschappelijk verantwoord inkopen” casus [The “socially responsible procurement” case].*

Kooijman, M., Schäfer, A., Stam, B., **Dewies, M.,** & Merkelbach, I. (2020). *Corona-proof Schiedamsedijk [Covid-19-proof Schiedamsedijk].*

Merkelbach, I., **Dewies, M.,** & Kooijman, M. (2020). *De “onnodig verkeersgeluid” casus [The “needless traffic noise” case].*

Merkelbach, I., **Dewies, M.,** Krijn, J., & Harkhoe, N. (2021). *De “no pyro, no party” casus [The “no pyro, no party” case].*

Merkelbach, I., Slag, G., **Dewies, M.,** & Kooijman, M. (2020). *Advies: stimuleren Corona-gedragaanwijzingen in het stadhuis [Advice: Compliance with Covid-19-related behavioural instructions in the city hall].*

Raven, L., **Dewies, M.,** Schop-Etman, A., Christiaanse, B., Boudaber, M., Brugts, J. & Zegveld-Steur, M. (2019). *De “De Wende” casus [The “De Wende” case].*

Schäfer, A., **Dewies, M.,** & Kooijman, M. (2020). *Stimuleren 1,5 meter afstand busstation Zuidplein [Stimulating 1.5 meters distancing at the Zuidplein bus station].*

Slag, G., **Dewies, M.,** Christiaanse, B., Schop-Etman, A., & Zegveld-Steur, M. (2020). *Contactrijk Betalen: Het stimuleren van betalingsafspraken naar aanleiding van een aangepaste nudgingbrief [Contact first, pay later: The stimulation of repayment by instalments as a result of an adapted nudging-letter].*

Stam, B., **Dewies, M.,** Janssen, A., Schop-Etman, A., & Zegveld-Steur, M. (2020). *De “concernpas” casus [The “employee badge” case].*

Stam, B., **Dewies, M.,** Schäfer, A., & Vogels, M. (2020). *De “slagboomduikers casus [The “traffic barrier divers” case].*

van den Boogaard, L., Kooijman, **M., Dewies, M.,** Merkelbach, I., Stam, B., van der Wielen, L., Slag, G., & Schäfer, A. (2020). *Advies heropening locaties concernhuisvesting [Advice reopening of work facilities].*

## Courses

Courses at the Erasmus Graduate School of Social Sciences and the Humanities.

Responsible research data management

Professionalism and integrity in research

Multilevel modelling II: Multilevel Structural Equation Modelling

Participatory action research

Photovoice

Qualitative data analysis

Data analysis with R

Mixed method research: How to combine diverse quantitative and qualitative methods

Doing ethnography

How to get your article published

Qualitative coding with ATLAS.ti

Qualitative interviewing

Qualitative comparative analysis

The focus group method

Courses from the German Academic Scholarship Foundation.

Vom Defizit zum Dialog: Einführung in die Wissenschaftskommunikation [From deficit to dialogue:

Introduction to science communication]

Wissenschaftskommunikation – Vortragen und Präsentation [Science communication: Lecturing and presentations]

I am thankful for the great support received from family, friends, and colleagues.



