

# Comprehensive Evaluation of the Behavioral Insights Group Rotterdam

Administration & Society  
2023, Vol. 55(8) 1555–1583  
© The Author(s) 2023



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/00953997231180302  
journals.sagepub.com/home/aas



Malte Dewies<sup>1</sup> , Inge Merkelbach<sup>1</sup>,  
Jurian Edelenbos<sup>1</sup>, Kirsten I. M. Rohde<sup>1</sup>, and  
Semiha Denктаş<sup>1</sup>

## Abstract

Behavioral insights teams (BITs) employ behavioral experts and policy professionals to collaboratively improve public policy. Most evaluations of BITs focus on the interventions that BITs develop, but not the functioning of BITs. Here, we report the first comprehensive evaluation of a BIT, the Behavioral Insights Group Rotterdam. We investigate how its resources were used, for what activities, with what outputs, and to which effects. Using quantitative and qualitative methods, we derive nine propositions to describe and improve the integration of behavioral insights into public policy and administration.

## Keywords

behavioral insights team, behavioral public administration, behavioral public policy, evaluation, nudge

---

<sup>1</sup>Erasmus University Rotterdam, The Netherlands

### Corresponding Author:

Malte Dewies, Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Burgemeester Oudlaan 50, Rotterdam 3062 PA, The Netherlands.  
Email: dewies@essb.eur.nl

The integration of behavioral science findings into public policy continues to attract widespread attention from governments and scientists alike (e.g., Afif et al., 2018; Grimmelikhuijsen et al., 2017; Lourenço et al., 2016; OECD, 2017). This integration, often generically referred to as behavioral insights, is based on the argument that public policy can be improved by avoiding common misconceptions about behavior (e.g., humans act like *homo economicus*) and infusing more realistic understandings of behavior into policy making and implementation (e.g., Thaler & Sunstein, 2008). Behavioral insight teams (BITs) are prominent forerunners translating this argument into practice. BITs combine expertise from behavioral sciences and public policy to address concrete policy issues with a behavioral dimension on a case-by-case basis by developing behavioral solutions. According to review articles, many solutions from BITs successfully change behavior (e.g., DellaVigna & Linos, 2020; Hummel & Maedche, 2019). Here, we report the first systematic evaluation of a BIT.

## Background

The publication of the book *Nudge* is a useful starting point for summarizing the evolution of behavioral insights (Whitehead et al., 2017). According to the book, nudges are light-touch interventions that tend to capitalize on behavioral automatisms to encourage behavioral change without forbidding any options (Thaler & Sunstein, 2008). *Nudge* became a worldwide best-seller, attracted widespread attention from public organizations, and paved the way for behavioral insights being institutionalized. The first BIT, also called “nudge unit” (Halpern, 2015), was formed in the United Kingdom (UK) only two years after *Nudge* was published. This team played an important role in combining behavioral insights with experimentation to trial mostly nudges in the field and find out “what works” (Haynes et al., 2012; John, 2014). Behavioral insights are since strongly associated with experimentation (Einfeld, 2019). According to Strassheim et al. (2015), behavioral insights gain authority using easily understandable experimental evidence, that “appeals to common sense reason, while at the same time being linked to scientific norms and standards.”

Not all BITs are the same, though and relevant differences were reported, for instance between BITs from Australia and New Zealand (Jones et al., 2021). However, BITs tend to have in common that they are specialized teams employing expertise from behavioral science and public policy for investigating policy issues from a behavioral perspective to develop solutions grounded in behavioral science findings. Many BITs rely on step-wise procedures to analytically identify interventions for changing behaviors underlying policy issues and on experimentation to evaluate those interventions before

implementation (OECD, 2019; Service et al., 2014). BITs also play an important role for disseminating behavioral insights by providing influential showcases and guidelines (e.g., Haynes et al., 2012; Service et al., 2014), and by capacity-building for policy professionals (e.g., Baggio et al., 2021).

Behavioral insights received relevant criticism. First, they were said to follow a limited understanding of autonomy and be part of an elitist program where government and/or scientists define good behavior and promote such behavior using a technocratic approach that suppresses policy debate (e.g., Feitsma, 2018). Second, nudges and similar efforts were claimed to be ineffective for tackling complex problems, and to distract from more durable system-level change (e.g., Selinger & Whyte, 2012). Third, the fixation on experimentation was claimed to limit the analytical lens of behavioral insights, the understanding of mechanisms, and the potential scope of application (e.g., Pearce & Raman, 2014). This criticism, however, motivated advanced versions of behavioral insights rather than preventing its growing popularity (Ewert, 2020; Ewert & Loer, 2021).

Key take-aways from these advanced versions are that behavioral insights should (a) draw on multidisciplinary inputs and engage various stakeholders to fully reflect the social and political embeddedness of behavior, (b) apply behavioral insights throughout all stages of the policy process, (c) integrate interventions into existing policy measures and aim to change social structures if needed, (d) embrace methodological pluralism, and (e) combine the micro-level focus on individual behavior with meso and macro level perspectives (Ewert, 2020; Ewert & Loer, 2021). In addition, the nudge concept was advanced to also stimulate deliberative decision-making processes (Banerjee & John, forthcoming) or boost individual decision-making competences (Reijula & Hertwig, 2022). Finally, scholars suggested that those involved in behavioral insights be aware and critical about behavioral factors influencing their own behavior (Lodge & Wegrich, 2016). The latter aspect relates to *behavioral public administration*, which is the “[. . .] analysis of public administration from the micro-level perspective of individual behavior and attitudes [. . .]” (Grimmelikhuijsen et al., 2017). In fact, advanced versions of behavioral insights widen the scope from behavior of the public to also include the behavior of administrators and interactions between both (Gofen et al., 2021).

## Present Study

What to the best of our knowledge is still lacking are systematic evaluations of BITs (McDavid & Henderson, 2021). In focusing mostly on the outputs of BITs (e.g., interventions; BIT, 2019; Lourenço et al., 2016), other features of BITs have received little attention (e.g., necessary

investments, effects on policy discourse, skill development; Kusters & Van Der Heijden, 2015). In an attempt to fill this gap, the current study reports the comprehensive evaluation of one BIT set up at a Dutch local government: the Behavioural Insights Group Rotterdam (BIG'R). BIG'R aimed to investigate how the municipal administration could benefit from behavioral insights as a public sector innovation lab (McGann et al., 2018). In line with that, the focus of BIG'R also was on the process of integrating behavioral science findings and public policy at the municipality, rather than developing interventions only. In adopting a broad focus and a long-term perspective, the evaluation covers a period of four years and investigates which resources BIG'R used, for what activities, with what outputs, and to which effects. It is a combined process and outcome evaluation.

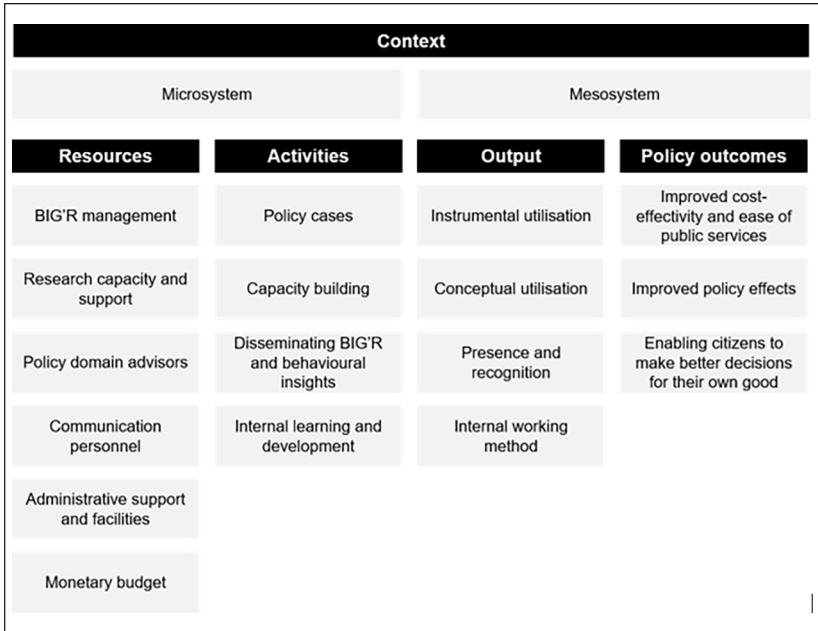
BIG'R was a "deviating case" (Gerring, 2008) when compared to other BITs (Ball & Head, 2021; Jones et al., 2021; Mukherjee & Giest, 2020; Sanders et al., 2018). For BIG'R, an academic institution and the municipality of Rotterdam acted as equal partners (a "boundary organization"; Guston, 2001), whereas for most other BITs either government or academia dominate. Moreover, BIG'R did not pre-commit to any behavioral or methodological framework for analyzing and changing behavior. Rather, it considered behavioral insights a collaborative and integrative effort between policy practitioners and researchers (Dewies et al., 2022). Finally, BIG'R decided to avoid political affiliations and support which is also different from many other BITs (e.g., John, 2014).

Scientific inquiry was the main motivation for the evaluation, but some results were also used to inform the municipality's decision to continue BIG'R in a different format (a utilization-focused evaluation; Patton, 1997). The research design has been reviewed by the DPECS ethics review committee at Erasmus University Rotterdam (20-023).

## **Methodology**

The evaluation was designed using a stepwise procedure adapted from Saunders et al. (2005) described below. We started the design after the first two years of BIG'R and data collection took place during the fourth and last year. The evaluation is hence predominantly summative in nature. Our aim was to derive propositions as key take-aways that can be tested in future research and support BIT practitioners for designing and developing BITs. Propositions are declarative statements about abstract constructs, here, based on empirical observations (i.e., induction).

We assume that our findings, although subject to biases, correspond to reality "out there" (i.e., post-positivist metatheory). To increase confidence in our



**Figure 1.** BIG'R logic model.

findings, we used a mixed methods design (Johnson & Onwuegbuzie, 2004) integrating multiple data sources (surveys, interviews, and documents) pragmatically (Morgan, 2007). Three of the authors (MD, IM, SD) had been members of BIG'R. These insider evaluators hold fine-grained experiences of BIG'R and interacted with informants naturally. To safeguard independence of the evaluation (Barnett & Camfield, 2016; Morris, 2004), two non-members (KR, JE) were involved in all stages of the research as well. We also pre-registered the research in advance to increase transparency (see <https://osf.io/f3av9>).

*Step 1: Logic Model*

To systematically describe BIG'R, we developed a logic model (Figure 1) depicting its relevant components and describing the underlying theory of change (Frechtling, 2007). The logic model was designed together with the BIG'R management through an iterative process of theorizing, feedback, discussion, and adjustment. As a starting point we took common basic elements of logic models (resources, activities, outputs, outcomes, and context; Frechtling, 2007) that we equipped with multiple blocks describing relevant

components of BIG'R. The logic model guided the evaluation and structures the presentation of our findings.

### *Step 2: Evaluation Questions*

For each block of the logic model, we developed evaluation questions. For resources, these questions investigated quantity and quality (e.g., number and education of BIG'R members), and for activities common aspects of process evaluations (fidelity, dose, reach, recruitment, responsiveness, and context; Durlak & DuPre, 2008; Saunders et al., 2005; Steckler & Linnan, 2002). For the other elements (output, policy outcomes, and context), the evaluation questions were directly inferred from the meaning of the respective blocks. This way, we were confident that the 98 evaluation questions we developed (see Supplemental Material) were sufficient for a comprehensive evaluation.

### *Step 3: Methods*

To answer all evaluation questions, each question was assigned to at least one of seven research methods (see Supplemental Material). Whenever possible, questions were assigned to multiple methods for triangulation. The seven methods were:

- (a) A randomized survey of a representative sample of municipality employees ( $N=779$ ).
- (b) Interviews with municipality employees who had collaborated with BIG'R for a policy case ( $N=28$ ).
- (c) Surveys of members of policy case teams ( $N=87$ ).
- (d) Surveys of instructors and attendees of capacity building activities ( $N=105$ ).
- (e) A survey of all BIG'R members ( $N=23$ ).
- (f) An analysis of documents (e.g., presentations and policy proposals).
- (g) A logbook kept by the first author.

### *Step 4: Interpretation and Integration*

To extrapolate from the highly granulated answers to the different evaluation questions (see Supplemental Material), we conducted three online sessions lasting 6 hours in total where one author (MD) presented the results to all co-authors who then wrote down their conclusions independently to limit the influence of groupthink (e.g., Park, 1990) and production blocking (Stroebe et al., 2010). Subsequently, the conclusions from all authors were discussed one-by-one with all authors during two additional online sessions of in total

4 hours to agree on relevant conclusions and propositions. In the presentation of our findings, we focus on these conclusions and propositions.

## Findings

In the following, we describe the findings for each block of the logic model.

### Context

*Microsystem.* The microsystem comprised the immediate physical and social work environment of BIG'R members. The physical environment included flexible workplaces at the office buildings of the municipality in the center of Rotterdam, one with office space exclusively available to BIG'R members. In addition, university employees could use workplaces at the university. During the last year of BIG'R, members mostly worked remotely because of the COVID-19 pandemic. Almost all BIG'R members had split work responsibilities (i.e., working part-time for BIG'R and part-time at other functions), meaning that their social work environment consisted of two or more separate layers.

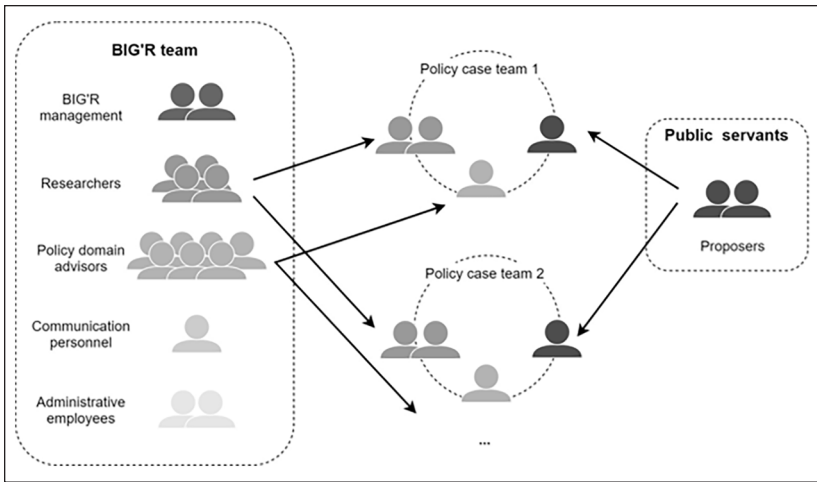
In general, the group climate within BIG'R was judged positively by BIG'R members. However, some members mentioned being kept out of information loops and reported misunderstandings. This may be due to the differing availabilities and work locations of BIG'R members, reducing opportunities for interaction. BIG'R members also acknowledged a difference in objectives between the university and the municipality (e.g., "doing things right" vs. "doing the right things," respectively).

*Mesosystem.* The mesosystem encompassed interrelations between the BIG'R microsystem and other microsystems, namely relations with other work settings. According to BIG'R members, having split work responsibilities could cause competing interests for available working time and priorities. However, some members also mentioned positive spillover effects (e.g., when behavioral insights were used for the other work environment).

### Resources

Resources were inputs required for the BIG'R activities. For role descriptions of the BIG'R management, researchers, and policy domain advisors, see Dewies and colleagues (Dewies et al., 2022) and Figure 2.

*BIG'R Management.* The BIG'R management consisted of a municipal project manager (0.44 FTE) and an academic head (0.20 FTE). Together, they



**Figure 2.** BIG'R team and composition of policy case teams.

were responsible for managerial tasks, such as management of people, the budget, team strategy, and output quality. The management itself and other BIG'R members found that management possessed sufficient knowledge and skills to fulfill its role.

Nevertheless, the management faced some challenges. First, management members had to get acquainted with the “other” organization (e.g., the municipal project manager learning the importance of scientific publications). Second, it reported that it invested more time than agreed upon and less than would have been optimal for developing the team and keeping a good overview of ongoing activities. Third, different skills and approaches were required during the start-up phase (i.e., focus on team formation) and later phases of BIG'R (i.e., focus on results). This refers to common group development processes that precede effective group performance (Tuckman & Jensen, 1977). To facilitate this process, another person took over the role of the municipal project manager after approximately two years.

**Research Capacity and Support.** BIG'R employed researchers from both the municipality (two 0.44 FTEs) and the university (1 FTE and 0.9 FTE) with diverse methodological and disciplinary backgrounds (e.g., psychology, anthropology). In general, researchers found that they possessed sufficient knowledge and skills to fulfill their role. However, they sometimes reported they lacked knowledge about the municipality administration as well as about behavioral (change) theories. According to the researchers, time constraints



could limit opportunities for reading (scientific) literature and could lead to preferring “quick fixes” over more careful working modes. Research capacity was perceived as an important bottle neck for policy cases; as a result, it was temporally increased using interns and student assistants.

It was sometimes unclear how the responsibilities between municipality and university researchers should be divided. In addition, there were some disagreements between both types of researchers about methodological choices, potentially implying enhanced scrutiny to combine scientific and municipal requirements. Uncertainties due to lack of clarity and disagreements were discussed and resolved via compromise, agreement, or management decisions.

*Domain Advisors.* The municipality administration was divided into seven subdivisions concerned with different policy domains (e.g., city development, public safety). Each subdivision seconded one policy domain advisor (PDA) to BIG'R (six 0.4 FTEs and one 0.2 FTE). About two-thirds of the PDAs reported that they invested more time, and about one-third less time than agreed upon. PDAs found that time constraints could limit efforts to disseminate knowledge about BIG'R and often resulted from part-time work duties that PDAs still had in their subdivisions.

Most PDAs held university degrees (in, e.g., criminology, history), and a few had received additional training related to behavioral insights. When compared to other BIG'R roles, more PDAs reported they lacked skills and knowledge to fulfil their role (e.g., lack of knowledge about the working method of BIG'R). Other BIG'R members found that availability and capacity differed between PDAs. A plausible contributing factor was a high turnover rate among PDAs, with a total of 17 different PDAs. Reasons for the high turnover were both unrelated and related to BIG'R (e.g., retirement, difficulty in combining part-time duties).

These results indicate that some PDAs found their role challenging. Adopting an interactional perspective on work experiences (Bakker & Demerouti, 2007), challenges originate from a lack of personal resources for too high task demands. The latter were high for PDAs because their role involved bridging the complex science-policy nexus (e.g., Strassheim, 2020a, 2020b) and, at the same time, learning about behavioral insights. Such bridging and learning required PDAs to understand the needs and relevance of science (i.e., having a positive attitude toward science and a basic understanding of research procedures and requirements) and to approach policy issues from a behavioral perspective (i.e., having a positive attitude toward applying behavioral insights, knowing how to apply behavioral insights). Without related resources (e.g., research training, a behavioral sciences background), this may have been too difficult, especially given the part-time

availability. The recruitment of PDAs did not take these resources into account as motivation and interest were the main criteria. These findings motivate our first proposition:

*P1: Competencies for conducting or managing research and knowledge about behavioral science are key for policy professionals working within BITs.*

This proposition corroborates earlier research (Feitsma & Schillemans, 2019; Jones et al., 2021). Such research also describes policy competencies and public sector experience as important for BIT professionals. PDAs, however, typically had many years of experience in the municipality in different roles, which may explain why this aspect did not feature in the evaluation.

**Communication Personnel.** One communication advisor was seconded to BIG'R from within the municipality (0.44 FTE). This person strategized and oversaw the group's external communication. Because of missing responses from communication personnel, most evaluation questions related to this resource cannot be answered. Other BIG'R members reported fluctuation in the available capacity of communication personnel, and on average they found communication personnel less sufficiently available than those in other roles.

**Administrative Support and Facilities.** Two administrative employees (0.44 FTEs) were responsible for operational aspects (e.g., monitoring the team's inbox). BIG'R members judged administrative employees to be sufficiently available. However, administrative employees themselves reported to have invested more time than agreed upon. They mostly held university degrees and did not miss any skills or knowledge to fulfil their assigned role according to both these employees and other BIG'R members.

**Monetary Budget.** The municipality funded BIG'R with an annual budget of 250,000 EUR. The budget was used to fund the university researchers (53% of the budget), personnel (36%), communication (8%), and research and organization (3%). BIG'R received additional indirect funding because many BIG'R members were paid by other units of the municipality. According to BIG'R management, no limitations were attributable to budget constraints.

## Activities

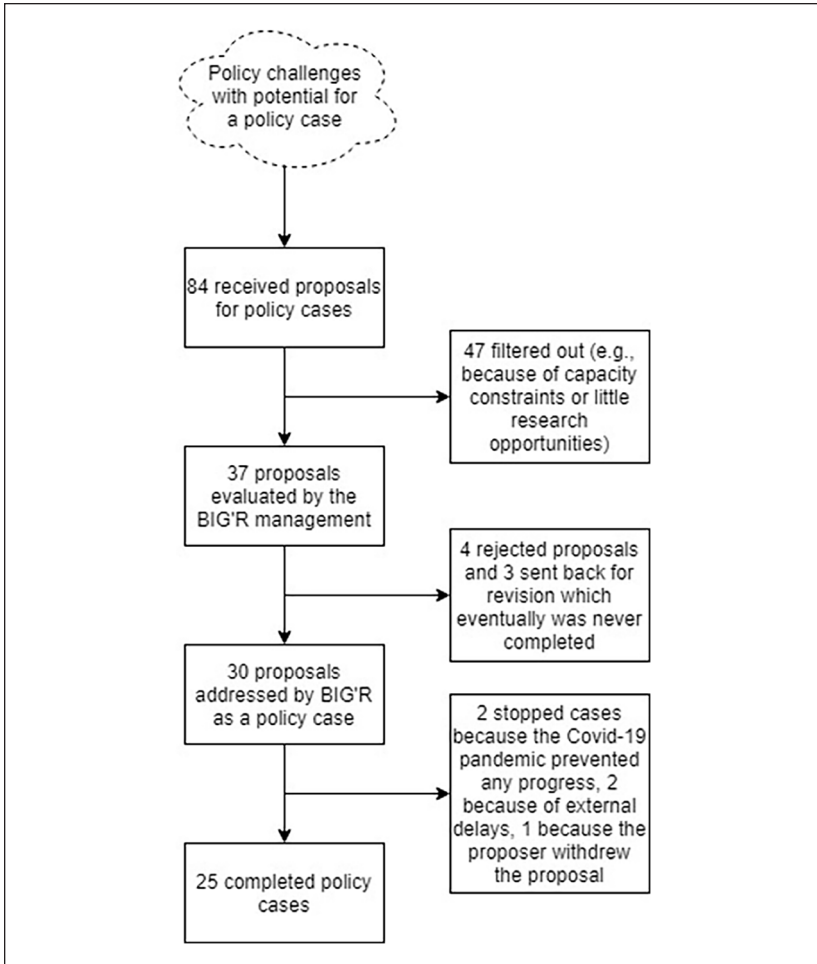
**Policy Cases.** Policy cases involved collaborative efforts of BIG'R researchers and PDAs to address concrete policy challenges of the municipality together

with municipality employees. Every municipality employee could propose a policy challenge related to the employee's work to be addressed together with BIG'R. In the following, we refer to these employees as proposers (Figure 2). Political actors were not allowed as proposers because BIG'R aimed to distance itself from party politics and because political actors were too remote from practice (e.g., they could not implement solutions or share field experience). According to PDAs, it was a challenge when political actors still approached BIG'R with suggestions for policy challenges because BIG'R then needed to find a suitable and motivated employee within the administration with whom to collaborate as a proposer.

Proposals for policy cases were stimulated by PDAs using their individual networks and communication channels (e.g., emailing department heads), and by dissemination activities that informed municipality employees about BIG'R. As a consequence, the number and nature of proposals could depend on individual PDAs (e.g., more proposals with more active PDAs). Most proposers worked on a strategic level (i.e., as project managers and policy advisors) or were communication staff. This may be due to a larger interest from these job functions as well as the set-up of BIG'R (e.g., PDA networks on a strategic level). Multiple reasons motivated proposers to submit a policy case. They often submitted policy challenges that they found relevant and urgent (e.g., related to political priorities), and many proposers wanted to try a new perspective for a wicked problem, hoping that a behavioral approach would lead to more effective solutions. In total, BIG'R received 84 proposals, 30 of which were addressed as a policy case and 25 were completed (Figure 3).

All policy cases followed a standard procedure of four phases: approval of the case, exploratory research to understand related behaviors, development of solutions, and efforts to stimulate implementation of policy advice from BIG'R. (For more details see Dewies et al., 2022.). During the first phase, policy case teams tightly defined target behaviors to be able to change and measure them. However, this could narrow the focus of policy cases (e.g., on residents in a specific neighborhood to recycle their food waste rather than Rotterdam residents to live more sustainably). In addition, most proposals started with a narrow focus on small parts of the municipality administration because of the fragmented administration (e.g., letters sent by one department rather than the whole administration).

Following the first phase, policy cases could take different forms to adapt to specific target behaviors and contexts. Policy case teams, for instance, used different forms of research to better understand behaviors, with the most common forms being desk research (93% of all policy cases), interviews



**Figure 3.** Flow of proposals for policy cases.

(38%) and site observations (31%). This required input from researchers with different perspectives and methodological backgrounds. For 43% of the policy cases, interventions were brainstormed with various stakeholders during a co-design session, and interventions were piloted in the field for 32% of the policy cases.

According to BIG'R members, involvement by proposers could differ between policy cases, with some proposers participating a good deal and

others demonstrating an outsourcing mentality (i.e., delegating the policy issue to BIG'R). Similarly, their time investment varied between 0.5 and 20 hours per month, according to the proposers. Some urgent policy cases, however, were completed quickly (e.g., within one month) while others required longer term collaboration (e.g., 2 years). The more involved proposers, for example, attended meetings, shared information and data, co-designed interventions and solutions, introduced BIG'R to stakeholders, and provided research assistance. After the collaboration, about half of the proposers reported that they continued their involvement with behavioral insights (e.g., reading related books).

The social climate within policy case teams was judged very positively for almost all cases. Multiple aspects of the collaboration (e.g., communication) and policy advice were judged positively as well. However, the high turnover of personnel that was already found to influence the availability of PDAs also impaired team composition and communication within policy case teams. In addition, the involvement of external stakeholders was a challenge in some policy cases when there was little engagement and support from those stakeholders, or their actions could not be foreseen. In general, BIG'R could act more like a facilitator or knowledge broker (Feitsma, 2019) when there were many stakeholders, and more like an independent problem-solver or policy-designer when there were only a few. This, and BIG'R's use of different research methods to investigate different kinds of target behaviors (e.g., on-off behaviors, habits, and group behavior) motivate the following proposition

*P2: BITs need to adapt their approach to policy cases (e.g., research methods, own role) to operate under different administrative circumstances and with different target behaviors.*

This proposition stands in contrast to the rule-based approaches propagated by other actors in the field that leave little room for adaptation (e.g., Kettle & Persian, 2022; OECD, 2019). These approaches typically encompass step-by-step procedures to develop and trial interventions, often anticipating a limited set of behavioral determinants. Meanwhile, our proposition emphasizes flexibility from BITs concerning procedures, research methods, and roles in interaction with stakeholders. This is in line with advanced versions of behavioral insights (Ewert, 2020).

**Capacity Building.** Capacity building encompassed 23 introductory trainings and few other activities of BIG'R members (e.g., sharing literature) to increase the knowledge and skills of municipality employees about

behavioral insights. Training sessions typically lasted about 1.5 hours and focused on how to brainstorm behavioral solutions for policy issues. Four hundred and four people attended the 16 training sessions, with attendance at individual sessions ranging from 7 to 48.

Most trainings were organized because BIG'R applied or was invited to present at events internal and external to the municipality. Training reached a similar target group as policy cases, since the majority of attendees worked on a strategic level and was motivated by curiosity about and openness to new ideas and tools. In addition, some attendees sought to use behavioral insights for specific applications.

Only four trainings took place after the start of this research and could be evaluated. In general, these sessions were judged positively (e.g., concerning the level of engagement) but the results are limited in their representativeness because they were conducted online during the Covid-19 pandemic. As trainings informed attendees about BIG'R and behavioral insights, trainings also were a dissemination activity.

*Disseminating BIG'R and Behavioral Insights.* Dissemination includes all efforts to inform municipality employees and others about BIG'R and behavioral insights. For this, BIG'R used its own corporate design and relied on various information channels (e.g., presentations, a website). In total, BIG'R members gave presentations at 89 events (including trainings), reaching 1,806 individuals during the 60 events for which the number of attendees was recorded. The reach for most digital products was low, however (e.g., a maximum of 79 readers in the municipality intranet).

On average, BIG'R members were not very content with dissemination (on average 3.7 on a scale from 1 "totally not content" to 5 "totally content"), mostly because they found communication not being up to date (e.g., an outdated website). The COVID-19 pandemic was considered a contextual factor that boosted dissemination because then behavior change was a necessity to increase compliance with hygiene measures, highlighting the importance of behavioral insights. In contrast, limited available time and communication expertise were contextual factors that hindered dissemination.

*Internal Learning and Development.* Internal learning encompassed capacity building for BIG'R members individually and as a group. This involved some formal training (e.g., on moderation techniques) but most learning happened on the job when discussing experiences from practice and lessons learned.

Generally, BIG'R members tended to be content with training and development (on average 4.0 on a scale from 1 "totally not content" to 5 "totally content"). Yet, they would have liked more learning and development

opportunities, and they found learning and development to be impaired by time constraints and a high turnover of personnel, resulting in little experience being accumulated. This suggests that high turnover not only complicated collaborations within policy case teams but also impaired learning and development. Therefore, we put forward the following proposition:

*P3: Team stability and/or good handover to new team members are key to the completion of policy cases, and to improve group learning and development.*

High turnover rates have been reported for some other BITs too (Fels, forthcoming; Jones et al., 2021), mostly in the context of internal staff mobility at their parent organizations. The Dutch government welcomed such mobility as a means of adaptation and renewal (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2015). While preventing risks from a high turnover, BITs can align with such a perspective and view inwards mobility as an in-flow of novel and diverse capabilities and outwards mobility as dissemination of knowledge and skills.

## Output

Outputs were the anticipated results of BIG'R activities. Borrowing from Pelz (1978), we differentiate between instrumental and conceptual utilization.

*Instrumental Utilization.* Instrumental utilization refers to the implementation of BIG'R policy advice. At least some aspects of advice were implemented for 15 of the 24 policy cases that were completed and where implementation could be investigated. For the other nine cases there was no implementation. Most implementations involved quick and simple changes (e.g., posters). Whenever BIG'R suggested more systemic or comprehensive solutions, proposers often were not supportive, because they found them too difficult to implement (e.g., getting support from stakeholders in other units of the municipality). Policy cases thus started with a focus on small parts of the administration related to the responsibilities of individual proposers (see Policy Cases section) and implementation was constrained when extending beyond these responsibilities. Therefore, we suggest the following proposition:

*P4: To increase their achieved scope of change, BITS require a broad mandate that includes support for implementation.*

Implementation challenges are rarely mentioned in the behavioral insights literature, possibly because experimental evidence is viewed as inarguable and the need to repackage and negotiate evidence for policy implementation is often downplayed (Einfeld, 2019). Our findings illustrate, however, that implementation is not a given and was limited by the resources and willingness of proposers who often needed to bridge different units of a fragmented administration for successful implementation. This highlights the importance of the micro level of individual behavior for policy implementation and challenges top-down views of mechanistic policy implementation (Gofen et al., 2021). Moreover, it highlights that BITs can benefit from being assigned a broad mandate from the start that encourages proposers to “think big” and guarantees access to resources needed for implementation (e.g., dedicated innovation teams). Otherwise, the scope of change achieved by BITs may be limited to “technocratic tweaks” (Hansen, 2018) that can easily be applied in different contexts and are well-known in the literature as “low-hanging fruits” (e.g., Sanders et al., 2018).

Case proposers gave multiple explanations for why advice was or was not implemented. The most common explanation for implementation was that proposers had co-produced the advice. Stimulating proposers to take an active role during policy cases can thus improve implementation. Moreover, for proposers, the advice often gained authority because it was co-authored by academic scholars referring to scientific evidence. This implies that BITs employing scholars can distance themselves from experimentation without immediately losing influence. We therefore suggest the following:

*P5: Besides direct forms of evidence (e.g., field experiments), BITs can also capitalize on circumstantial forms of evidence (e.g., expert authority) to gain influence.*

As mentioned in the introduction, BITs often gain authority by referring to experimentation that is presented as the highest form of evidence (Einfeld, 2019; Feitsma, 2020). If BITs can refer to circumstantial forms of evidence instead, this enables BITs to supply and advocate behavioral insights that cannot be prototyped and changed for experimentation (e.g., policy directions). This way, BITs may more easily achieve change that is not just incremental (e.g., Halpern & Mason, 2015). It moreover enables BITs to better integrate different sources of evidence (e.g., meta-analyses) instead of having to rely on singular experiments.

Interestingly, some proposers adopted a behavioral problem understanding during the policy case trajectory (e.g., some cars are too loud *because* drivers want to attract attention) that required a behavioral



solution (e.g., create better opportunities to attract attention) rather than standard approaches (e.g., policing tuned cars). Such an understanding could facilitate implementation and was plausibly fostered when proposers were more actively involved, providing them with more instances to acquire this understanding. We propose the following:

*P6: Explaining the mechanisms underlying behavioral solutions to decision makers leads to better implementation.*

Comparing the factors that we found to influence implementation with factors from a comprehensive literature review (Damschroder et al., 2009), we find behavioral problem understanding to be a novel mechanism for implementation. However, behavioral problem understanding can also be an outcome, suggesting that behavioral insights can be applied during all stages of policy making, particularly problem definition (Ewert, 2020; Gopalan & Priog, 2016). In the past, behavioral insights often were viewed as a means to develop interventions rather than an analytical lens.

Common explanations for not implementing advice were a lack of someone coordinating and pushing implementation, lack of urgency of the policy issue, the requirement to involve multiple stakeholders, and time and capacity constraints. In addition, interventions not fitting their targeted context was a frequent reason for not implementing them (e.g., a recommendation to install streets signs was not implemented because legal enforcement of the signs was impossible). This implies that intervention development should anticipate the intervention's target context and individuals involved in their implementation, since decision-making about implementation is dispersed across strategic and operational levels (Mintzberg & Waters, 1985; Tummers & Bekkers, 2014). Since there was no automatic implementation mechanism, support for implementation from other parts of the municipality administration had to be secured actively by BIG'R members and proposers. We therefore suggest the following:

*P7: For better implementation, BITs can benefit from more concentrated implementation efforts and insights from implementation science.*

With the behavioral insights literature often assuming that evidence translates into implementation, there has been little uptake of knowledge from implementation science. Implementation science, however, points to factors that increase the likelihood of advice being implemented (e.g., Fixsen et al., 2005), suggesting that BITs can, in fact, plan for implementation (e.g., Dewies et al., 2022).

**Conceptual Utilization.** Conceptual utilization refers to what proposers learned about integrating behavioral insights into public policy. Generally, lessons learned fell in three broad categories. The largest category encompassed aspects directly related to behavioral insights, namely the direct application of behavioral insights to public policy (e.g., by using social norms), an increased awareness of behavioral aspects and their importance for public policy, and a focus on the target group and its behavior. Proposers with a background unrelated to social or behavioral sciences sometimes described the collaboration as an “eye-opener” for the possibility to program behavior rather than taking behavior as a given. The second category concerned lessons learned about collaboration, namely that expectation management and process management are important. The third category encompassed lessons learned about science, namely that science can be impractical but also useful in providing convincing evidence and a neutral, task-oriented perspective.

*P8: Policy professionals can deliberately collaborate with BITs to learn about the importance and usefulness of behavioral insights for public policy.*

We believe conceptual utilization covers relevant learning outcomes affecting the attitudes and skills of proposers. We find this aspect to be lacking in the literature, since most research focuses on members of BITs rather than their collaborators (e.g., Ball, 2022). Proposers can thus be viewed as learners rather than clients or informants for intervention development. This perspective can help to view and design collaborations with BITs as a curriculum-based learning experience (Billett, 2011) that enables proposers to apply behavioral insights independently and BITs to engage in applied ways of capacity-building.

**Presence and Recognition.** Presence and recognition refer to behavioral insights and BIG'R being known and positively appreciated. After about 3.5 years of BIG'R, one third of municipality employees indicated they know the group. In their view, the group had a neutral (e.g., meeting minimum requirements) or no reputation (e.g., not having formed an opinion) since the average score was close to the middle of the scale (3.1 on a scale from 1 “negative” to 5 “positive”), with some variation between policy subdivisions (range 2.50 – 3.88).

**Internal Working Method.** The internal working method refers to a written document that described a standardized way of collaborating for BIG'R members, designed to instruct policy case teams. The document was regularly reviewed and adjusted. According to BIG'R members, major adjustments included the addition of more concentrated efforts to achieve

implementation of policy advice, to hold more role-specific instead of general meetings, and to clarify role responsibilities. Generally, these changes represent a shift toward a more result-oriented working mode. In addition, after recognizing that not every policy case could be combined with scientific research that pilots interventions, BIG'R also approved policy cases in which it advised solutions without prior piloting and adjusted the working method accordingly. BIG'R members tended to be somewhat content with the final working method (on average 3.6 on a scale from 1 "totally not content" to 5 "totally content"). However, some considered it still work in progress, and some recognized a difference between the document and practice (e.g., some members did not always comply with the working method).

### *Policy Outcomes*

Policy outcomes refer to the consequences of policy cases for municipal practices and how they relate to three goals that were defined for BIG'R at its start (Figure 1). For instance, one policy case helped to achieve better cost-effectiveness of municipal services by reducing clean-up costs for garbage (Merkelbach et al., 2021). According to proposers, some but not all policy cases contributed to achieving the goals of BIG'R (Table 1). If they did, they typically contributed to achieving some but not all of the goals at the same time. One obvious reason for policy cases not contributing to reaching the goals was that advice from BIG'R was not implemented.

### **Discussion**

We report, to the best of our knowledge, the first comprehensive evaluation of a BIT. BITs rely on expertise from policy and behavioral sciences to integrate behavioral science findings and public policy. We evaluated BIG'R which was the BIT of the municipality of Rotterdam for a period of 4 years. A logic model served to systematically describe BIG'R and allowed us to evaluate all its relevant aspects. Our findings led us to the formulation of eight propositions, to which we will add one more below.

The municipality administration was fragmented horizontally into different subdivisions, and vertically based on multiple levels separating strategic from operational tasks. In theory, BIG'R aimed to integrate behavioral insights across the whole administration but in practice the subdivisions with more active PDAs and strategic levels of the administration were better reached. We believe, however, that operational levels, although not reached well by BIG'R, have much to gain from behavioral insights because the literature reports many related examples (e.g., BIT, 2019). Fragmentation also

**Table 1.** BIG'R Contributions to Policy Outcomes.*Goal*

## BIG'R contribution

*Improved cost-effectiveness*

- + Garbage collection/cleaning costs were reduced because residents put their garbage in the dedicated containers rather than disposing it in the streets.
- + Policing costs that resulted from incorrectly parked bicycles were reduced because more bicycles were parked in dedicated parking spaces.
- + Canvassing costs for offering help to residents with financial problems were reduced because a letter giving notice of the canvassing encouraged residents to seek help proactively or cancel canvassing.
- + Potential intervention costs to stop residents hanging out with friends in the lobby of a public swimming pool were prevented.
- + Investment costs in an online app unlikely to be effective in stimulating sustainable behavior were avoided.
- Costs for increasing compliance with hygiene measures were higher because more communication material was produced.
- Extra time investments were made because of the collaboration with BIG'R

*Improved ease of using public services*

- + Being aware of hygiene measures at municipal service centers eased compliance.
- + Compliance with hygiene measures was eased at municipality offices because more and better opportunities were created to comply.
- + Walking routes were more accessible because bicycles were parked correctly, not blocking the way.
- + Understanding of a letter offering help for finding a job was eased.
- + Correctly understanding that the entry of a public swimming pool was to welcome residents rather than to hang out was eased.
- + Stakeholders involved in equipping a transport hub to make travelers comply with hygiene measures got input and advice from the municipality.
- Using public services may not become inherently easier, but the desired behavior may appear more attractive (i.e., reducing felt unease).

*Improved policy effects*

- + Health risks were reduced for visitors of municipal service centers because compliance with hygiene measures was improved.
- + Health risks were reduced for municipality employees and visitors of municipality offices because compliance with hygiene measures was improved.
- + Cleaner and more pleasant neighborhoods were created because less garbage was disposed in the streets.
- + Simpler communication with unemployed residents who got offered help for finding a job was achieved.

*(continued)*

**Table 1. (continued)**

Goal
BIG'R contribution
<ul style="list-style-type: none"> <li>+ Football supporters were given a voice and they were involved for finding solutions to containing the use of illegal fireworks in stadiums during matches.</li> </ul> <p><i>Enabling citizens to make better choices</i></p> <ul style="list-style-type: none"> <li>+ Choices to comply with hygiene measures were improved because residents and employees of the municipality were more aware of these measures and how to comply with them.</li> <li>+ Automatic processes were triggered that caused residents to dispose their garbage correctly in dedicated containers rather than in the streets, without deliberately thinking about it.</li> <li>+ Residents with financial problems were better informed about opportunities to receive help with financial problems, thereby enabling them to make a deliberate choice about whether or not using this help.</li> <li>+ Residents were encouraged to park their bicycle correctly, enabling others to pass by faster.</li> <li>+ Discouraging hanging out at the entry of a public swimming pool created a more welcoming, pleasant, and peaceful environment for users of that swimming pool.</li> </ul> <ul style="list-style-type: none"> <li>• Defining what is the better choice may be paternalistic.</li> </ul>

Note. Plus (minus) signs indicate how BIG'R contributed to (not) reaching the goals; reflective sidenotes from interviewees do not have a sign.

•indicates reflective sidenotes from interviewees.

meant that policy case teams rarely operated in isolation but were dependent on various stakeholders (vertically and horizontally) for the successful completion of case studies and the implementation of solutions. This motivates the final proposition.

*P9: Effective boundary workers are important to increase reach and complete policy cases.*

BITs are often “policy labs” (McGann et al., 2018) that are separated from other parts of the organization to be able to come up with novel solutions. However, this increases the distance between BITs and the organization, requiring “boundary workers” (e.g., Langley et al., 2019) to bridge this distance. BITs can reflect on this bridging (e.g., where should behavioral insights be integrated, who should be involved in intervention development and how) to link their activities to their goals. Such reflections seem central to the success of BITs but scholarship has rarely addressed.

BIG'R followed an analytic approach and extensively investigated policy issues that it addressed in order to improve understanding of these issues. Extensively investigating issues served to develop contextualised understanding and to incorporate different sources of knowledge and disciplinary inputs (Dewies et al., 2022). This enabled BIG'R to provide advice that did not develop stand-alone concepts but sought to complement existing policy. This differs from BITs that repeatedly exploit a limited set of behavioral automatisms to develop interventions. Yet, despite recent scholarly advice (e.g., Ewert, 2020), BIG'R involved those targeted by interventions (e.g., residents) only in some instances. Therefore, it may be that BIG'R sometimes promoted behaviors wanted by municipality employees and researchers rather than "helping people make the choices they want to make" (White, 2013, p. 101). It is plausible that the involvement of an academic institution and collaborations with BIG'R being free of charge were contributing factors in motivating and enabling enhanced efforts of BIG'R to improve understanding of policy issues before rushing to solutions.

Scholars criticised the fixation of behavioral insights on the micro level of individual behavior while neglecting the meso and macro levels (e.g., Ewert, 2020; Moseley & Thomann, 2021). The logic model that we developed and our findings respond to this criticism by describing what is needed at the meso-level to change behaviors at the micro level. It links individual policy cases and behavioral change attempts with administrative aspects and policy processes, acknowledging that the application of behavioral insights needs to be strategized and embedded in policy practice. This enables others to reflect on BITs more comprehensively and shift attention from interventions developed by BITs to how BITs operate, how they generate what outputs, and how they are embedded in policy practice. Our propositions are first steps investigating contingencies between different aspects of BITs (e.g., how do resources influence activities and outputs of BITs?) to improve BIT practice and its effects.

This evaluation looked at actors within the municipal administration using a behavioral lens, for instance, investigating their motivations and attitudes. In this way, the evaluation contributes to the field of behavioral public administration (Grimmelikhuijsen et al., 2017): First, it corroborates the argument that implementation of policies (e.g., advice from BIG'R) is contingent on the perceptions, understandings, and willingness of administrators (Moseley & Thomann, 2021). Second, the propositions provide advice for the staffing and management of personnel in BITs. Third, the evaluation illustrates how BITs can be leveraged and embedded within organizations to improve the skills and knowledge about behavior in those organizations.

Like all studies, this evaluation has limitations. Specifically, the double role that some of us had as evaluators and BIG'R members might have caused some biases. For instance, positive results may reflect having asked the wrong questions rather than signs of good program implementation and outcomes. However, the non-members of BIG'R involved in this research helped to reduce such bias. Second, as a "deviant case" BIG'R differs from many other BITs, limiting the potential for generalization (Mullin, 2021). BIG'R's distinctiveness, however, suggests that this evaluation concerns novel aspects of BITs that might be of broader interest or possible application. Finally, data collection for this evaluation started only during the last year of BIG'R, producing some delay between events and data collection. Therefore, accurately recalling original experiences might have been more difficult due to decaying or distorted memories.

Future research can apply and evaluate our propositions in the field. In addition, future research can use our findings to compare the design of different BITs and what changes they bring about. Such research can identify best practices for the set-up and operations of BITs. Finally, future research can use government employees who collaborated with BITs as informants. They were major informants for this research, providing critical comments based on their experience that related, for instance, to ethics and the limits of behavioral insights.

Taken together, BIG'R incorporated most aspects of advanced versions of behavioral insights. Yet, it encountered challenges that received limited scholarly attention so far, particularly insufficient implementation, the need to employ effective boundary workers, and aspects of embedding and strategizing the application of behavioral insights within organizations.

### **Authors' Note**

Malte Dewies is also affiliated with University of Cambridge, United Kingdom.

### **Acknowledgments**

We thank the municipality of Rotterdam for this research opportunity. We thank Astrid Schop-Etman for her support in designing the research.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**

Malte Dewies  <https://orcid.org/0000-0002-0576-2085>

**Supplemental Material**

Supplemental material for this article is available online.

**References**

- 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>
- 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. (2022). Translating behavioral public policy into practice: Interpretations and traditions. *Critical Policy Studies*, 16(3), 315–332. <https://doi.org/10.1080/19460171.2021.1973525>
- 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>
- Banerjee, S., & John, P. (Forthcoming). Nudge plus: Incorporating reflection into behavioral public policy. *Behavioural Public Policy*. <https://doi.org/10.1017/bpp.2021.6>
- Barnett, C., & Camfield, L. (2016). Ethics in evaluation. *Journal of Development Effectiveness*, 8(4), 528–534. <https://doi.org/10.1080/19439342.2016.1244554>
- Behavioural Insights Team. (2019). *Annual report 2017-18*.
- Billett, S. (2011). Workplace curriculum: Practice and propositions. In F. Dochy, D. Gijbels, M. Segers, & P. van den Bossche (Eds.), *Theories of learning for the workplace: Building blocks for training and professional development programs* (pp. 17–36). Routledge.
- 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>
- 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>
- Dewies, M., Denктаş, 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>
- 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>
- 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>
- 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>
- 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. (2019). 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. (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>
- Fels, K. M. (Forthcoming). Who nudges whom? Expert opinions on behavioural field experiments with public partners. *Behavioural Public Policy*. <https://doi.org/10.1017/bpp.2022.14>
- Fixsen, D. L., Naom, 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. <https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf>
- Frechtling, J. A. (2007). *Logic modeling methods in program evaluation*. Jossey-Bass.
- Gerring, J. (2008). Case selection for case-study analysis: Qualitative and quantitative techniques. In J. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford handbook of political methodology* (pp. 645–684). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199286546.003.0028>
- 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>

- Gopalan, M., & Pirog, M. A. (2017). Applying behavioral insights in policy analysis: Recent trends in the united states. *Policy Studies Journal*, 45(S1), 82–114. <https://doi.org/10.1111/psj.12202>
- 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>
- 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>
- Halpern, D. (2015). *Inside the nudge unit: How small changes can make a big difference*. WH Allen.
- Halpern, D., & Mason, D. (2015). Radical incrementalism. *Evaluation*, 21(2), 143–149. <https://doi.org/10.1177/1356389015578895>
- Hansen, P. G. (2018). What are we forgetting? *Behavioural Public Policy*, 2(2), 190–197. <https://doi.org/10.1017/bpp.2018.13>
- Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). *Test, learn, adapt: Developing public policy with randomised controlled trials*. Behavioural Insights Team. <https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials>
- 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>
- 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 of Psychology*, 12, 660410. <https://doi.org/10.3389/fpsyg.2021.660410>
- 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>
- 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>
- 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>
- Kettle, S., & Persian, R. (2022). *Target, explore, solution, trial, scale: An introduction to running simple behavioural insights projects*. Behavioural Insights Team. <https://www.bi.team/wp-content/uploads/2022/02/BIT85-Running-behavioural-insight-trials-report-v4.pdf>
- 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>

- 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>
- 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>
- 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>
- 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>
- McGann, M., Blomkamp, 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>
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2015). *Werken is bewegen: Arbeidsmobiliteit in het openbaar bestuur*. [https://kennisopenbaarbestuur.nl/media/207856/Werken\\_is\\_bewegen\\_arbeidsmobiliteit\\_in\\_het\\_openbaar\\_bestuur.pdf](https://kennisopenbaarbestuur.nl/media/207856/Werken_is_bewegen_arbeidsmobiliteit_in_het_openbaar_bestuur.pdf)
- 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>
- 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>
- Moseley, A., & Thomann, E. (2021). A behavioural model of heuristics and biases in frontline policy implementation. *Policy & Politics*, 49(1), 49–67. <https://doi.org/10.1332/030557320X15967973532891>
- 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>
- Mullin, M. (2021). Learning from local government research partnerships in a fragmented political setting. *Public Administration Review*, 81(5), 978–982. <https://doi.org/10.1111/puar.13395>
- OECD. (2017). *Behavioural insights and public policy: Lessons from around the world*. <https://doi.org/10.1787/9789264270480-en>
- OECD. (2019). *Tools and ethics for applied behavioural insights: The BASIC toolkit*. <https://doi.org/10.1787/9ea76a8f-en>
- 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.
- 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.
- Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy, 6*(1), 119–149. <https://doi.org/10.1017/bpp.2020.5>
- 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>
- 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.
- Steckler, A. B., & Linnan, L. (2002). *Process evaluation for public health interventions and research*. Jossey-Bass.
- 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., 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)
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- 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), 527–547. <https://doi.org/10.1080/14719037.2013.841978>
- White, M. D. (2013). *The manipulation of choice: Ethics and libertarian paternalism*. Palgrave Macmillan. <https://doi.org/10.1057/9781137313577>
- Whitehead, M., Jones, R., Lilley, R., Pykett, J., & Howell, R. (2017). *Neoliberalism: Behavioural government in the twenty-first century*. Routledge. <https://doi.org/10.4324/9781315684772>

### Author Biographies

**Malte Dewies** is research associate at the El-Erian Institute of Behavioural Economics and Policy at the University of Cambridge, United Kingdom. Prior to this position, Malte obtained his PhD at Erasmus University Rotterdam, the Netherlands. His research interest is the intersection of psychological science and public policy.

**Inge Merkelbach** is assistant professor of Behavioral Change and Societal Transitions at Erasmus University Rotterdam, and managing director of the Behavioural Insights Group Rotterdam (BIG'R), a behavioral scientific impact center. She focusses on contributing to the solution of complex societal issues by applying behavioral insights in practice.

**Jurian Edelenbos** is professor in the field of Interactive Urban Governance at Erasmus University Rotterdam. His research interests are community-based initiatives, trust in interorganizational collaboration, participatory processes, and boundary spanning leadership.

**Kirsten I. M. Rohde** is professor of Behavioral Economics at Erasmus University Rotterdam. Her main research interests are time, social, and risk preferences. In her research she uses a combination of decision theoretical and empirical approaches.

**Semiha Denктаş** is professor of Psychology at Erasmus University Rotterdam. Her research interests are health psychology, inclusion and diversity, and applied psychology.