

**Nijmegen School of Management**  
**Department of Economics and Business Economics**  
**Master's Thesis Economics (MAN-MTHEC)**

# **From insights to implementation: a study on the adoption of behavioural insights in Dutch policy**

By Bente Hogema (s1028674)  
Nijmegen, 29 June 2024

Program: Master's Program in Economics  
Specialisation: Economics, Behaviour and Policy  
Supervisor: Koen van der Swaluw

**Radboud Universiteit**



**Abstract**

There is a world-wide growth of institutions that focus on applying behavioural economics to policy. After the example of DellaVigna et al. (2022), this study examines the adoption of behavioural insights, but with the scope on Dutch policies. The study consists of two legs, one of quantitative nature to determine the degree of adoption and one of qualitative nature, to disclose the underlying mechanisms. Using the projects published by Behavioural Insights Network Netherlands (BIN NL), it is determined that in approximately 50% of the cases, findings of behavioural research are used in varying extents in policy initiatives. If an intervention is proposed by an individual with a policy-related background, it is more likely to get adopted compared to when it is proposed by a behavioural advisor. Organizational factors, the political context of policy development and scalability issues form (among others) barriers for implementing behavioural insights. It is found that turnover in the workplace complicates the ability to hold on to interventions, whereas commitment within the organization facilitates adoption of behavioural insights in policy.

## Table of content

1	Introduction.....	4
1.1	Research question .....	5
1.2	Thesis outline .....	6
2	Literature review .....	7
2.1	Behavioral economics .....	7
2.1.1	Behavioral economics history in a nutshell .....	8
2.1.2	The rise of behavioral insights in practice .....	9
2.1.3	The Dutch behavioral-institutional landscape .....	9
2.2	Integrating behavioral insights and public policy .....	10
2.2.1	When to apply behavioral insights.....	10
2.3	Issues with implementing behavioral insights .....	11
2.3.1	Overreliance on RCTs.....	11
2.3.2	Threats to good scientific practice .....	12
2.3.3	Limited understanding of context .....	12
2.3.4	Bounded rationality of policy professionals .....	12
2.3.5	Scalability .....	13
2.3.6	Fleeting nature of behavioral insights.....	13
2.4	Hypotheses .....	14
3	Data and methods.....	17
3.1	Quantitative data .....	17
3.1.1	Sample and survey .....	17
3.1.2	Features of projects .....	18
3.1.3	Adoption .....	19
3.1.4	Methodology .....	20
3.1.5	AI verification.....	21
3.2	Qualitative data .....	21
3.2.1	Sample selection .....	22
3.2.2	Interview content .....	23
3.2.3	Methodology .....	24
3.2.4	Preparing the data .....	24
3.2.5	Coding.....	24

---

4	Results from quantitative analysis .....	26
4.1	Representativeness of the sample .....	26
4.2	Average adoption .....	26
4.3	Predictor: evidence.....	28
4.4	Predictor: type of organization.....	29
4.5	Predictor: type of intervention .....	30
4.6	Predictors in a logistic regression model.....	32
4.7	Open-answer questions .....	34
4.7.1	AI verification.....	34
5	Results from qualitative analysis .....	36
5.1	Representativeness .....	36
5.2	General barriers of adoption .....	37
5.2.1	Additional barriers .....	39
5.3	Barriers of adoption – in case of effective interventions .....	40
5.3.1	Longevity of interventions.....	41
5.4	Success factors .....	42
6	Discussion and conclusion.....	44
6.1	Conclusion .....	44
6.1.1	Extent of policy adoption.....	44
6.1.2	Explaining policy adoption.....	45
6.2	Discussion .....	46
6.3	Strengths & limitations .....	48
6.4	Research recommendations .....	49
6.5	Policy recommendations.....	51
7	References.....	53
	Appendix I – Interview guide.....	59
	Appendix II – Overview of open codes.....	60
	Appendix III – Code tree .....	61
	Appendix IV – Robustness checks .....	62
	Appendix V – Categorization of psychological biases.....	64

## 1 Introduction

The popularity of behavioral economics has surged throughout the last decade, seemingly provoking a shift in how governments choose to address policy issues (van Bavel, 2020). Equipped with the task to continuously decide which policies to implement and thus determine what approaches are both most effective and efficient (Hoogerwerf, 2014), what better premise could governments wish for than the one of behavioral economics? After all, this field of science, emerging from the combination of both economics and psychology, should facilitate to create policy tools that are considerably more effective than the traditional tools (Madrian, 2014).

The popularity of behavioral economics is illustrated by a world-wide growth in institutions that focus on the application of behavioral insights to public policy. Ultimately, it requires the right policy capacity to substitute behavioral findings into policy making (Kuehnhanss, 2019). The first behavioral insights units were established in 2010, whereas the total number of institutions engaged in this topic reached over 300 in 2023 (OPSI, 2023). With the realm of behavioral economics expanding, the Netherlands too, is engaged in improving policy based on behavioral insights. To help effectuate this, Behavioral Insights Network Nederland (BIN NL) was founded in 2014. BIN NL is a collaboration between Dutch ministries and government agencies, whose main goal is to ‘share behavioral insights in the community, with the aim of bringing the application of behavioral knowledge further’ (BIN NL, n.d.). In addition to the founding of BIN NL, the governmental interest in behavioral insights is reflected by the Behavioral Insights Teams (BITs) within ministries, or in some case municipalities that focus on developing more behavior-oriented policies.

The considerable growth of behavioral insights units arguably contributes to a larger reflection of these insights in policies. BIN NL confirms this to some extent in the Netherlands: ‘since 2014, behavioral insights are gradually but increasingly utilized in policy, execution, supervision and communication (BIN NL, 2022). What is mentioned additionally however, is that the utilization happens on an ad hoc basis and that a cohesive and structural embedding in the policy process is lacking. Even though there appears to be an increase in adoption, data on the exact uptake of behavioral insights are missing.

When looking at a global scale, the evidence of adoption is limited as well. This can firstly be attributed to the fact that the actual adoption of these insights is not inherently evident: Wang and Yang (2021) argue that the complexities of the environment that policies are

developed within are likely to affect gathered insights. Adding to this paper, DellaVigna, Kim and Linos (2022) investigate how often innovations from Behavioral Insights Team North America (BIT-NA) were adopted in policy and what factors moderate the adoption. They find that because of bottlenecks mainly related to organizational inertia, Randomized Controlled Trials (RCTs) ‘only yield about one third of their potential benefits’ (DellaVigna et al., 2022). The second explanation as to why there is lacking evidence of adoption, is because the current body of research that examines the uptake of behavioral innovations is limited.

On the one hand, there is the ‘promise’ of enhancing policy by integrating behavioral science techniques, which is reflected by a global growth in research institutions focused on behavioral insights. On the other hand, the quantitative findings on this topic are limited and a perusal of evidence adoption in policies the Netherlands is lacking. A study aimed at examining the current situation of the uptake of behavioral insights in policy can therefore be justified.

## **1.1 Research question**

The goal of this study is to examine the extent to which research of behavioral insights is implemented in policies in the Netherlands. Building upon that, it goes into the possible barriers policymakers face when taking up innovations in policies. Gaining a better understanding of the integration of evidence in policy making consequently allows for improving the uptake. The research question this paper seeks to answer is: To what extent have the findings of behavioral insights projects been integrated in policy initiatives in the Netherlands? This question is answered using data of behavioral insights trials conducted throughout the years 2014 - 2024, as compiled in the BIN NL database. A second research question that ensues from this topic is: What bottlenecks do policymakers face when adopting behavioral insights in policy? This question is aimed at examining the experiences of policymakers to yield a more comprehensive grasp of the underlying mechanisms.

The contribution of this research is twofold. The scientific relevance lies in adding to the existing body of literature regarding the uptake of behavioral insights in policy. As previously mentioned, the current body of literature is limited and findings of the few relevant studies conducted are not necessarily applicable to the Netherlands. Indeed, DellaVigna et al. (2022) name the context specificity of their paper as a large limitation. This study seeks to both identify and validate predictors of evidence adoption in policy in the Dutch context, and this way give direction to the scientific debate. By doing so, it contributes to decreasing the current knowledge gap. The societal relevance of this study is reflected by suggestions for improving

the application of behavioral insights in policy. Disclosing the precise mechanisms at play is required in order to address occurring bottlenecks. Examining possible factors that foster or complicate this process is therefore relevant for both researchers and policymakers. A better notion of what prevents policy professionals from adapting research findings can help design practical strategies to address these issues. This way, the eventual application of behavioral insights to policy in the Netherlands can be improved.

## **1.2 Thesis outline**

Chapter 2 of this paper elaborates on relevant theories and this way provides a theoretical framework as a foundation for the hypotheses. As mentioned before, the current literature on the uptake of evidence in policy is limited. This means that rather than providing a summary of evidence adoption on a global scale, the chapter gives an overview of relevant theories to create an intuition of the mechanisms behind the uptake of innovations. Chapter 3 addresses the methodology and the process of data collection. The quantitative results are presented and analyzed in Chapter 4, whereas Chapter 5 covers the results of the qualitative research. The paper concludes with Chapter 6, which discusses the implications and limitations of the study and provides a final conclusion.

## 2 Literature review

This chapter defines the most important concepts and illustrates the gap between economics and public administration, and its implications for policy development. Before considering the present-day implications of behavioral insights for policy, a brief overview of the merging of the disciplines of economics and psychology is given. Because previous results of a systematic uptake are yet lacking, a broader overview of the policy development process is provided, which allows to recognize where exactly in the process possible issues with the uptake may arise. The literature presents different frameworks and guidelines on the application of behavioral insights to policy. Next, this chapter provides an overview of the main takeaways. Following from the literature discussed, hypotheses for this study are drafted.

### 2.1 Behavioral economics

Before assessing the uptake of behavioral insights in policy, a definition of both the concept behavioral insights as well as policy is required. The term behavioral insights is used to some degree interchangeably with behavioral economics. This can be explained by the main takeaways of these concepts – from a public administration point of view at least - arguably narrowing down to the same idea, that is integrating an improved and evidence-based understanding of human behavior to the policy making process (Troussard & van Bavel, 2018). A distinction between these concepts can however be made. There is no consensus on a uniform definition of the term behavioral economics, but in general they refer to the scientific discipline that combines economics, psychology and other social sciences, which comes hand in hand with the violation of the assumptions of standard economic theory (Lunn, 2014). Behavioral economics relaxes the traditional assumptions of the rational agent model, since this model does not describe people well enough (Kahneman, 2011). As Thaler (2016) phrased it: ‘behavioral economics replaces the Homo economicus (Econs) with Homo sapiens (Humans)’.

Behavioral insights can be seen as building blocks for behavioral economics. The concept refers to ‘pieces of knowledge based on empirical findings about behavior’ (Troussard & van Bavel, 2018). These pieces of knowledge are generally expressions of the intuitive and unconscious thinking process we use for making quick decisions, rather than the deliberate process used for making conscious choices (Kahneman, 2011). Applying behavioral insights to policy would not make much sense if humans were rational to the extent that traditional economics assumes. Altering aspects of the choice architecture for example and other changes



of this nature would not affect the choice outcomes if humans were perfectly rational (Congdon & Shankar, 2018).

Behavioral economics also deviates from standard economic theory when it comes to the scientific methodology: behavioral economics is about applying scientific approaches of other sciences to economics (Shiller, 2005). The standard economic approach starts out with a theory based on assumptions, which is then tested empirically. Contradicting this deductive approach, behavioral economics infers its theory from experiments and systematic observations (Lunn, 2012; 2014). The rooting in empirics makes behavioral economics an inductive science, opposed to the deductive nature of standard economics.

### **2.1.1 Behavioral economics history in a nutshell**

If behavioral economics is about human behavior, then what exactly is non-behavioral economics about? By means of this question, Thaler (2016) contends that behavioral economics appears to be a pleonasm. After all, economics is a social science about the acting and interacting of humans (McChesney, 2013). Integrating the behavioral perspective in economics has however not always been as evident as it is in present-day. Economics as a field emerged before psychology did and during the 20<sup>th</sup> century it was attempted – for example by Milton Friedman - to shelter the economic assumptions from psychological considerations (Sent, 2004). This more psychological approach was deemed necessary in the first place because of an issue arising with standard economic theory: it aims to not only characterize optimal behavior, but also predict actual behavior (Thaler, 2016). From a historical perspective, behavioral economics can be divided into two streams, that is ‘old’ and ‘new’ behavioral economics (Heukelom & Sent, 2017).

The old behavioral economics is a movement that emerged halfway through the 20<sup>th</sup> century as a result of a collective discontent with mainstream economics. It aimed to develop an alternative that would include psychological insights (Sent, 2004). This movement included Herbert Simon and George Katona as ambassadors, but was unable to create a lasting impact, partly because of the attempts to distance itself from mainstream economics (Heukelom & Sent, 2017).

New behavioral economics arose in the late 1970s and listed Daniel Kahneman and Amos Tversky as best known ambassadors. The starting point of the new behavioral movement was mainstream economics and it examined deviations from this stream, opposed to developing an entirely new philosophy (Heukelom & Sent, 2017; Truc, 2021). The popularity of behavioral

economics grew massively during the years 1990 up till now, to a degree that all contemporary economics can - to some extent - be considered behavioral economics now (Agner, 2019).

### **2.1.2 The rise of behavioral insights in practice**

The increasing prevalence of behavioral economics is not only evident in literature (Truc, 2021; Nagatsu, 2015), but also in public policy delivery. The increase in attention towards this science is often associated with the books ‘Nudge’ (written by Thaler & Sunstein, 2008) and ‘Predictably irrational’ (written by Ariely, 2008), which acted as a catalyst in the process (Feitsma, 2019a; Raymaekers & Brans, 2020). Whitehead et al. (2014) find that behavioral sciences have influenced the public policy delivery in some form, in at least 135 states. Given the popularity of the science, it is not unlikely that this number has increased even further in the last decade. The attention for behavioral insights in policy in the Netherlands gained traction with the publishing of two rapports by the Scientific Counsel for Government Policy (WRR) in 2014<sup>1</sup> and 2017<sup>2</sup>. Following the latter rapport, the ‘capacity to act test’ was established by the cabinet in place then, obligating to take into account individuals capacity to act in policy development (*Kamerstukken II*, 34775, nr. 88, 2017/18).

### **2.1.3 The Dutch behavioral-institutional landscape**

Behavioral insights in an institutional sense are incorporated in different ways in the Dutch administrative system. With regards of using behavioral insights in policy, the local governments have the advantage of being organized closer to citizens and can affect their environment or how they behave in this environment. Nonetheless, on this level, the application of behavioral insights happens in a fragmented and diffuse way (Feitsma, 2019b). In the Dutch central government, behavioral expertise is organized partly according to the BIT set-up in the United Kingdom: there are five (small, consisting of a maximum of five people) BITs active, throughout different policy areas. In addition to the BITs there are other, less formal behavioral initiatives in the shape of ‘networks, research programs and working groups’. Regardless of the increasing interest in behavioral insights in the administrative system, a fundamental inclusion of the principles in the institutional sense is lacking (Feitsma, 2019a).

To promote using behavioral insights and facilitate behavioral pilot programs throughout all ministries, BIN NL was founded in 2014 as a collaboration between ministries and other

<sup>1</sup> Policymaking using behavioral expertise, WRR 2014

<sup>2</sup> Why knowing what to do is not enough. A realistic perspective on self-reliance, WRR 2017

public services. This way, the Dutch government gives substance to its intentions of ‘encouraging the application of behavioral insights in policymaking.’ BIN NL has a role in enhancing the exchange of knowledge and experiences and supports the application of behavioral insights in policy, implementation and supervision within the government. BIN NL reports on the progress of using behavioral insights in policy every two years by bundling and publishing behavioral insights projects (A Wealth of Behavioral Insights). In addition to these (and other publications), BIN NL hosts an annual Behavioral Insights Day (BIN NL, n.d.).

## **2.2 Integrating behavioral insights and public policy**

As stated before, the purpose of using behavioral insights in public policy is that it can create policy tools that are more effective and efficient compared to the traditional policy tools. Whether or not behavioral insights are used in policy largely depends on how one defines policy. Again, this is a concept without an universal definition. This study maintains the definition of policy provided by Hoogerwerf (2014) that states that policy is ‘the pursuit of certain objectives through certain means within certain temporal choices.’

### **2.2.1 When to apply behavioral insights**

The use of economics as a source for policy development is not new. In fact, economics as a science has been of great relevance for informing policy (Ahluwalia, 2015; Alm, 2017). The question of when to apply behavioral insights to policy refers to both the condition of a policy itself and at what point exactly in the policy process. Whenever there is a behavioral component in policy, behavioral insights can be applied. These behavioral components can be present in different ways (Bavel et al., 2013). The most obvious example is the case in which behavioral change is the main purpose of the policy, for instance getting people to exercise more to attain a healthier lifestyle. Secondly, there is a behavioral component to policy when people’s behavioral reactions affect the effectiveness of a certain policy. This is for example apparent in the case of consumer protection. Finally, there are behavioral components to the policy making process itself, in the shape of people working in the administrative system, who are subject to heuristics and biases themselves. Integrating behavioral insights in policy is thus not limited to a certain policy domain. Instead, they are applicable in a wide range of policy areas, varying from health and consumer protection to sustainability and taxation (Thaler & Sunstein, 2008; Lourenço et al., 2016).

Behavioral insights can be applied to public policy at different phases of the policy process. The policy cycle is a simplification used to conceptualize and organize this process (Capano & Pritoni, 2020). The cycle is both descriptive and prescriptive of nature, meaning it provides an outline of the different stages of the development of policies as well as a guideline for how to navigate through these stages (Cairney, 2012). The policy cycle is commonly divided into five stages, that is agenda setting, policy formulation, legitimation, implementation and evaluation.

Applying behavioral insights to policy can broadly speaking occur in three different ways, depending on the phase of the policy cycle the policy making process is in. Firstly, at the very start of policy development. Using behavioral insights during the primary stages of policy development can help to get a better grasp of the problem, which in turn leads to more tailored policy tools. Furthermore, behavioral insights can also be adopted in the implementation phase, meaning that the insights are being taken into account when designing the policy instrument. Finally, they can be used for tweaking policies that are already being implemented. This entails that the insights are being used directly to alter people's behaviour, in the phase where the policy is being applied (Troussard & van Bavel, 2018). Regardless of the variety of options available for integrating behavioral insights in policy, they are now mostly used to refine policy that is already developed (Van de Riet, Pietersma & Enserink, 2023).

## **2.3 Issues with implementing behavioral insights**

The positive implications of behavioral insights on policy are often highlighted: it can offer new policy tools, can create a better indication of the effects of already existing policies and it can help generate new welfare implications (Benartzi et al., 2017; Szaszi et al., 2017; Chetty, 2015). It should be noted however that there are several narratives that both counterweigh the mentioned advantages and complicate the adoption of behavioral insights. These challenges are captured in the framework of Dewies et al. (2022) and comprise four different topics, respectively an overreliance on RCTs, threats to good scientific practice, a limited understanding of context and bounded rationality of professionals applying behavioral insights. This section illustrates the framework and suggests two additional propositions.

### **2.3.1 Overreliance on RCTs**

For the purpose of pinpointing the exact effects of an intervention, behavioral insights are commonly studied in a – to a large extent - isolated environment, usually in the shape of a RCT

(OECD, 2018). Using RCTs as a research method for behavioral insights is promoted and done often enough for it to become ‘the golden standard’ in behavioral insights research (Ramos & Matos, 2018). As a consequence, BITs tend to refrain from doing any research that is not suitable for RCTs (Ball & Head, 2021) which creates a bias in the body of behavioral insights research. Basing evidence-based policies solely on the evidence gathered in RCTs excludes other sources of evidence policymakers could benefit from (Feitsma, 2019a).

### **2.3.2 Threats to good scientific practice**

In line with this, Dewies et al. (2022) find that threats to good scientific practice are another complicating factor in implementing behavioral insights. These threats occur because behavioral insights research generally takes place in a governmental realm, which can lead to compromised methodological designs and interferences in the shape of a change in the political tide or the turnover of public servants (Bolton & Newell, 2017). Along the lines of this argument, Lichand and Serdeira (2023) find that one of the reasons behavioral insights units are not ‘thriving’ as promised, is partly because of cherry picking in the academic world. Behavioral units worldwide do not yield the same success with their projects behavioral interventions as academic studies do. This can be attributed to (amongst others) an overrepresentation of good results in academic journals, which paint a brighter picture of behavioral insights than reality holds.

### **2.3.3 Limited understanding of context**

Furthermore, Dewies et al. (2022) state that behavioral insights research tends to overlook the context that human behavior is inherently embedded in. The context referred to is for example one’s (early) childhood or the institutional environment people are located in. As a result of this simplification of reality, the policies designed might not be as effective as expected.

### **2.3.4 Bounded rationality of policy professionals**

The fourth and final issue Dewies et al. (2022) describe concerning the implementation of behavioral insights in public policy, is the bounded rationality of policy professionals. Policy professionals – regardless of their aim to develop impartial policies - are themselves susceptible to decision making fallacies (Banuri, Dercon & Gauri, 2017). This is also referred to as the rationality paradox: ‘an approach that places bounded rationality at the heart of its thinking reflects little about the limits of its own rationality’ (Lodge & Wegrich, 2016).

Moreover, the domains of economics and politics are inevitably connected when adopting behavioral insights in policies. A policy process that concerns governmental policy is part of the political cycle that takes place between the political system on the one hand and the societal environment on the other hand (Hoogerwerf, 2014). In general it is found that the use of behavioral insights in policy making hinges largely on the interpretations of the research findings, and how they are transformed into the policy arena (Kuehnhanss, 2019). Evidence coming hand in hand with ‘emotional appeals’ can be seen as a way to manipulate the policy agenda and can outweigh other governance principles (Cairney & Oliver, 2017; Cairney, 2019). Indeed, Vivalt and Coville (2022) emphasize that due to the biases policymakers might have when interpreting evidence, the uptake of behavioral findings does not automatically lead to improved policies. Since the process of policy making is inherently rooted in a political context, the evidence of innovations is inevitably altered by the agendas and biases of involved stakeholders and can therefore not edict policies (Head, 2015).

### **2.3.5 Scalability**

An additional challenge that comes hand in hand with adopting behavioral insights in policy is the issue of scalability. This issue is twofold. Firstly, even when interventions work, they are only being scaled up a handful of times, for example due to lacking resources (Lichand & Serdeira, 2023). The second issue concerns the effectiveness of interventions when they are being scaled. RCTs are often used for pinpointing the exact effects of an intervention (OECD, 2018). The results of an experiment might however be valid in the specific context of the trial, but do not necessarily yield the same effects on a larger scale. The process of the cost-benefit ratio increasing - and thus the loss of efficacy - when extrapolating an idea from a smaller to a larger scale is referred to as a ‘voltage-drop’ or ‘voltage effect’. Behavioral insights trials usually come in the shape of RCTs or experiments and are hence the textbook example of something that ‘works in a petri dish but falls apart at scale’ (List, 2024; Brandon et al., 2022; Al-Ubaydli et al., 2020). Interventions might thus look promising at first glance, but lose their intended effects when used at any scale other than the pilot where they started.

### **2.3.6 Fleeting nature of behavioral insights**

Another caveat that presents itself when integrating behavioral insights in policy making is that the effects of behavioral insights can be fleeting by nature. Thaler (2018) provides an example of the default setting losing its effect when people get immediate feedback and find out that the option is not to their liking. This way, the intervention has little to no effect. This fleeting nature

of interventions can be illustrated by an example of a project in the BIN NL database as well. A study aimed at increasing safety for cyclists shows that the initial effect of the intervention has disappeared completely in a follow-up measurements two months after the intervention (placing a board sign) - arguably due to the lasting exposure (BIN NL, 2023).

## 2.4 Hypotheses

The determinants of adoption of behavioral insights in policy used in this research are along the lines of DellaVigna et al. (2022) and extrapolated to the Dutch context. DellaVigna et al. (2022) distinguish three categories of determinants. First of all they describe the statistical evidence. Secondly they recognize the features of the organization, in their research thus referring to characteristics of a city. The third category encompasses the experimental design, that is the type of treatment or the mechanism within the nudge. For this paper, the determinants of adoption are translated into the Dutch context. The formal hypotheses that are tested in Chapter 4 are provided underneath each paragraph.

The first category of predictors, that is strength of evidence, encompasses whether or not the project found a positive and statistically significant effect size of the intervention used. It is expected that a significant effect - indicating that the intervention in fact increases the display of desired behavior - contributes positively to the uptake of the project in policy. In addition to this, is tested whether the use of a control group in a project can predict the adoption of the project in policy. The use of a control group allows to attribute the change in behavior more concisely to the intervention, which positively affects the robustness. As a result, the findings of the project could be perceived as more convincing by policymakers and thus be more likely to be adopted.

*H0<sub>a</sub> – There is no effect of effect size on adoption*

*H1<sub>a</sub> – There is a positive effect of effect size on adoption*

*H0<sub>b</sub> – There is no effect of use of a control group on adoption*

*H1<sub>b</sub> – There is a positive effect of use of a control group on adoption*

The second category of predictors, that is organizational features, refers to the characteristics of institutions involved. One of the dimensions within this category is state capacity, which is proxied by the size of governmental organization involved in the project. Following the argumentation of DellaVigna et al. (2022), the larger the organization involved, the more likely it is that an innovation is adopted.

*H0<sub>c</sub> – There is no effect of the organizational features on adoption*

*H1c – There is a positive effect of the organizational features on adoption*

The final category of predictors concerns the type of innovation used in the project. DellaVigna et al. (2022) distinguish two groups of interventions, that is simplification and social comparison or personal motivation. The distinguishment between these two groups happens based on how uncontroversial or how aggressive the intervention is deemed, where simplification is considered the first and social comparison or personal motivation the latter. It is argued that projects considered less invasive are more likely to be implemented than interventions deemed more aggressive. A second dimension in this category is whether or not the intervention in the project was pre-existing. An example of a pre-existing intervention would be if the intervention is sending a letter, and this letter was sent before the project as well but now tweaked by the intervention, it is labelled pre-existing. If the sending of a letter did not happen before the intervention, it is considered new communication. It can be argued that pre-existing communication corresponds with a higher rate of adoption, because of lower implementation costs of the intervention.

*H0d – There is no effect of intervention type on adoption*

*H1d – There is a positive effect of intervention type on adoption*

*H0e – There is no effect of pre-existing communication on adoption*

*H1e – There is a positive effect of pre-existing communication on adoption*

The dimensions of the three categories are presented above are embodied in the conceptual model in Figure 1.

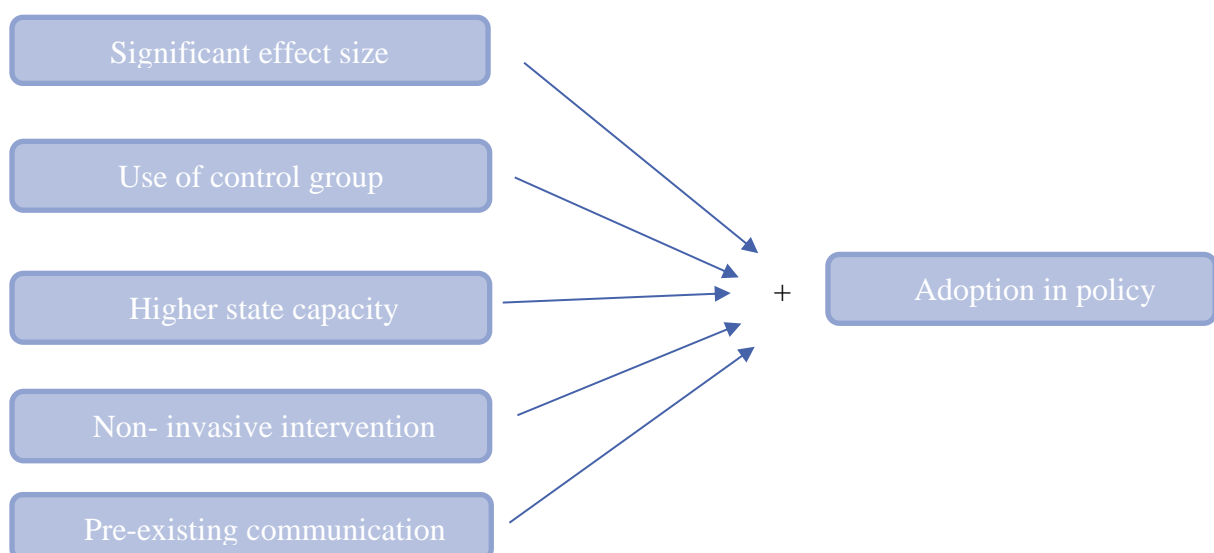


FIGURE 1: CONCEPTUAL MODEL



In addition to the mentioned categories, a new predictor is added, capturing the position of the one who initiated the project. This predictor is added to the model to help understand the influence of this position. Since systematic evidence on the uptake of behavioral insights in policy is as yet missing, there is little work to base hypotheses besides DellaVigna et al. (2022). This new predictor therefore is of explorative nature. It could be argued that if someone in a power position initiates the project, it will be more likely to get adopted. This makes sense, considering implementing the project would be in their interest and they are able to exert a certain amount of control in decision making. While a relationship between the initiator of the project and the uptake of innovations makes sense theoretically, there is thus far no validation.

*H<sub>0f</sub> – There is no effect of initiator of the project on adoption*

*H<sub>1f</sub> – There is a positive effect of initiator of the project on adoption*

### 3 Data and methods

This chapter provides an overview of the data and the methodology used in the study. It covers the quantitative branch of the research first, describing the process of data collection and analysis. Subsequently, it presents the qualitative side of the paper, elaborating on the methods employed for gathering and interpreting the interview data.

#### 3.1 Quantitative data

This study investigates the uptake of behavioral insights in policy in the Netherlands. To narrow down the scope of the research to a more manageable size, the focus is on the behavioral trials that have been published by BIN NL and the extent to which these findings were used in policy after the trials were finished. These researches were conducted by different parties, varying from ministries to governmental organizations and municipalities. BIN NL does not play a role in conducting research itself.

The common denominator of these projects is that they all examine a behavioral intervention. The projects are published on the BIN NL website, in the BIN NL database. Between the years 2014 and 2024 there was a total of 108 publications<sup>3</sup> of behavioral interventions in the database (BIN NL, n.d). The projects cover a wide range of themes, varying from sustainable behavior to finance, and from cybersecurity to student loans. The aim of the projects is either to increase a certain behavior, such as getting a certain vaccination, or to decrease a certain behavior, such as placing garbage next to a container on the streets. The vast majority of these interventions has a similar research set-up. They all comprise a behavioral analysis, in which the effects on behavior are measured before the intervention and in a follow-up measurement. Part of the projects also used a randomly selected control group, that did not receive the treatment (BIN NL, 2023).

##### 3.1.1 Sample and survey

The initial sample consists of the 108 projects that were published in the BIN NL database between 2014 and 2024. The BIN NL database contains sporadic information on what eventually happens with the intervention in terms of policy integration after the trials are completed. To obtain clear insight into if and how exactly the results of the projects are used,

<sup>3</sup> 108 projects was the number at the start of this study. During the research, at least one extra project was published in the database that is not taken into account in the analysis in this paper.

the lead authors were asked about this in a Qualtrics survey. The survey was drafted by example of the survey used in DellaVigna et al. (2022) and the main purpose was to establish the eventual uptake of the projects. The survey also covered the initial goal of the project (whether it was aimed to land in policy or not) and the representativeness of the project for the job of the respondent. The respondents were asked for a prediction of the rate of adoption for the 108 projects, they were asked about what they believe to be reasons for non-adoption and finally, they were asked what happened with the research results in terms of policy adoption. Using a survey to obtain these insights was considered most effective given the timeframe in which data collection had to take place. Nonetheless, this method did have implications for the final sample selection, because if the contact person of the research was not willing to fill in the survey, the project could not be included in the final sample. To keep the barrier for participating as low as possible the survey was composed in Dutch. Some of the prediction questions originally proposed in DellaVigna et al. (2022) were left out to minimize the required time to fill in the questions.

The authors were approached by their email address, provided with their research in the BIN NL database. When there was no email address provided, the author was contacted via LinkedIn or the cooperating organization was approached, either on the general email address or the contact form provided on the website. Two reminding emails were sent after the initial invitation, and the survey was brought to the attention of (intended) respondents by BIN NL (Thomas Dirkmaat), also per email. For the final sample selection, all the projects belonging to authors who did not complete the survey were removed. Non-completion happened in cases where the author either did not respond to the invite of filling in the survey or only completed the survey for part of their submitted research. In most cases of non-completion there was simple no response from the approached researchers. In four cases it was made known that the reason for non-completion was that the author had found a new job. Since it was not possible to establish if and how findings of projects were used in policy when the survey was not completed, the projects of non-completed surveys were removed from the sample. This ultimately resulted in a sample size of 46 projects.

### **3.1.2 Features of projects**

The projects in the total set of projects were scored in an Excel file on features presented in Chapter 2, including the significance of the effect size, the use of a control group in the experiment and the type of experimental design, i.e. the type of intervention used in the trial

(for example anchoring; social norms; loss aversion etc.). From this file, a second data file was distilled, containing only the projects from the final sample. In addition to the previously scored features, specific features derived from the surveys were added to the projects, among which degree of adoption and the role of the person initiating the project. All these features serve as predictor variables in the analysis. The predictors and their possible outcome values are presented in Table 1.

<b>Predictor</b>	<b>Outcome values</b>
Adoption – <i>whether or not findings of a project were used in policy</i>	Yes/No
Aim of project – <i>whether or not the initial aim of a project was to affect policy</i>	Yes/No
Effect size – <i>whether or not a significant positive effect was found in a project</i>	Yes/No
Control group – <i>whether or not a control group was used in the project</i>	Yes/No
Organization type – <i>covering the type of organization involved in the project</i>	Ministry; other type of (governmental) organization; combination of a ministry and other type of organization
Intervention type – <i>whether or not an intervention regarded as controversial is used in the project</i>	Yes/No
Pre-existing communication – <i>whether or not the communication of an intervention already existed in some shape</i>	Yes/No
Initiator of the project – <i>covering the work-related position of the individual initiating the behavioral project</i>	Behavioral advisor or BIT; policy officer; combination of BA/BIT and policy officer

TABLE 1: PREDICTOR VARIABLES AND THE CORRESPONDING OUTCOME VALUES

### 3.1.3 Adoption

Adoption is defined based on the definition provided by DellaVigna et al. (2022), referring to the case in which a behavioral technique used in policy after the research. As described before, the policy process consists of different phases. Adoption includes the adoption of behavioral insights in all phases of the policy process. In other words, a BIN NL project does not necessarily have to lead to the development of an entire new policy idea in order to be labeled

adopted. The tweaking of already existing policies and adjustments in the implementation as a result of the project is also regarded as adoption. Respondents were asked about the extent of adoption of the project in the survey. The options for the extent of adoption respondents could choose from not adopted, moderately adopted, largely adopted or completely adopted. Respondents could also indicate they did not know if the findings of a project had eventual implications for policy. Regardless of the degree of adoption, all the projects that were to some degree adopted, were finally labeled as adopted. Projects of which it was unknown if the findings were adopted were labeled as not adopted. This results in a binary dependent variable (Yes/No)

### 3.1.4 Methodology

A binary logistic regression model is used for the quantitative analysis. This means that rather than predicting a value for adoption, the probability of adoption occurring is calculated, based on the predictors used (Field, 2018). This type of analysis was chosen because both the fit with the collected data and its ‘ability to produce a predictive representation of real world situations’ (Healy, 2006). The following model captures the logistic regression:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n)}}$$

Where  $P(Y)$  represents the probability of adoption occurring,  $e$  is the base of natural logarithms,  $b_0$  represents the intercept of the standard linear model,  $b_i$  covers the relationship between each predictor and its outcome and  $x_i$  represent the value of the predictor variables. Next to this logistic regression model, a two-proportion Z-test is used to calculate differences between percentages of adoption associated with certain features of groups.

The conventional alpha level in statistical analysis is set at .05 (Cowles & Davis, 1982). Indeed, Keuzenkamp and Magnus (1995) state that ‘If economists have natural constants, then the most well-known is 0.05.’ More recently however, concerns have been raised about using this conventional level of significance and the implications it has for the occurrence of Type I errors, or in other words for replicating results. Using an alpha of 0.005 is a proposed alternative to avoid reporting false positives (Benjamin et al., 2018). Following this trend in research it would make sense to set the alpha level at a relatively small probability, i.e. 0.01 or 0.005. Nonetheless, there are other factors that affect the choice of the significance level in research (Miller & Ulrich, 2019; Kim & Choi, 2021) and for choosing the alpha level in this study, both the sample size and the nature of the research should be deliberated. The study examines 46 observations, which is a relatively small sample size. A smaller sample size comes hand in

hand with not only larger variability of the data, but also with decreased statistical power, which makes it harder to detect small effect sizes (Haile, 2023). In addition to this it should be noted that in studies of an explorative kind or pilot studies, it is not uncommon to use a higher alpha level, varying from .1 to .2 to prevent type II errors (Serdar et al., 2021). This also applies to this study, since the aim is to detect a possible relationship between the predictor variables and the possible adoption of evidence in policy. Because of the limited sample size and the explorative nature of this study, an alpha level of 0.05 was chosen in the statistical analysis.

### **3.1.5 AI verification**

In the survey was one open-ended question included, asking the respondents what they believed to be the main reason for non-adoption of the findings of the BIN NL projects in policy. The answers were read and interpreted, and the main themes emerging are summarized in section 4.7. After these topics were described, the answers to the open-ended question as filled in in Qualtrics were entered into ChatGPT as an explorative type of validation. ChatGPT was asked '*what are the main themes in the this data?*'. The output ChatGPT generated is presented in section 4.7.1.

## **3.2 Qualitative data**

In order to get a better understanding of the reasons behind the possible uptake of behavioral insights in policy, 14 structured interviews were conducted. Interviews are regarded best-suited for this study, as they allow to ascertain and theorize issues (Jamshed, 2014) regarding evidence adoption in policy. The interviews provide a more comprehensive illustration and can encompass nuances a survey does not cover.

The interviews were held in Dutch in cooperation with another master student from Radboud University, whose study topic to some extent aligned with the topic of this study. The interviews were conducted online, as a digital setting allowed for easier scheduling for everyone involved, in the short timeframe available. The majority of the interviews took approximately one hour, with the exceptions of an interviewee who was short in time (limited to approximately half an hour) and an interviewee who was willing to engage a bit longer (approximately one and a half hour).

### 3.2.1 Sample selection

The sample of interviewees consisted of behavioral experts on the one hand, and policy professionals on the other hand. The group of behavioral advisors is formed by people working at Behavioral Insights Teams from different Dutch ministries, behavioral researchers for Dutch municipalities and behavioral researchers for executive bodies. The group of policy advisors is composed of people working at Dutch ministries or Dutch executive bodies. The final sample thus includes both behavioral advisors and individuals who are concerned with developing policies in their daily jobs. This specific sample selection is deemed not only knowledgeable about the topic of interest, but can also provide insights from different point of views. The sample of interviewees is hence composed in line with purposeful sampling (Palinkas et al., 2013).

Similar to approaching the sample of the surveys, the interviewees were reached out to by the e-mail address provided with a behavioral insight project they worked on that is published by BIN NL. For the first selection of interviewees, individuals that collaborated on multiple projects published by BIN NL were approached. This resulted in 8 interviews, of which 7 were conducted with one interviewee and one was conducted in a dual setting (i.e. with two interviewees). By asking the interviewees in the interviews directly if they knew possible other suited participants for the interviews, ('the snowball strategy' (Creswell, 2016)), two additional individuals were contacted, which resulted in two additional interviews. In order to get in touch with policy professionals suited for the interviews, emails were sent out to behavioral advisors who cooperated in interviews, in which they were asked for references. This resulted in three additional interviews with policy advisors. An overview of all the respondents and their references is provided in Table 2 below. This specification of a reference number is used to link quotes to individuals in the results section.

<b>Interviewee reference</b>	<b>Role</b>	<b>Organization</b>
R1	Behavioral advisor	Municipality
R2	Behavioral advisor	Central government
R3	Behavioral advisor	Central government
R4	Behavioral advisor	Central government
R5	Behavioral advisor	Central government

R6	Behavioral advisor	Central government
R7	Behavioral advisor	Central government
R8	Behavioral advisor	Central government
R9	Behavioral advisor	Executive body
R10	Behavioral advisor	Central government
R11	Behavioral advisor	Executive body
R12	Behavioral advisor	Central government
R13	Policy advisor	Executive body
R14	Policy advisor	Executive body
R15	Policy advisor	Central Government

TABLE 2: OVERVIEW OF RESPONDENTS

### 3.2.2 Interview content

The interviews cover three themes. Firstly, the representatives of projects published by BIN NL for behavioral insight projects in general. Respondents who cooperated on these projects were asked to what extent they believed the impact of BIN NL published projects to represent their projects that were not published.

Secondly, the interviewees were asked about barriers they encounter with implementing behavioral insights in general. A distinguishment between ‘general’ barriers and barriers occurring specifically when implementing a tested intervention was made. For the first type of barriers, respondents were asked if they recognized certain barriers in their work, for example a lack of time or a lack of stimuli from the manager. The barriers respondents are asked about originate from the study of the Dutch Organization for Applied Scientific Research (TNO) in the Directorate-General of Climate and Energy in 2023 (Paradies et al., 2023).

For the latter, which the emphasis of this study is on, respondents were not asked if they recognized barriers, but they were encouraged to talk about their own experiences. The decision to list examples of ‘general’ barriers and ask about the extent of recognition was made for the purpose of saving time in the interviews. The aim of the study after all, is to explore possible explanations that go beyond what is already known.

The third theme covered in the interviews are the factors that positively contribute to adoption of behavioral insights in policy. Interviewees were asked what in their experience are success factors for the uptake of behavioral insights, and how the current uptake could be improved. The interview guide can be found in Appendix I. Given the volume of the behavioral



insights teams operating in the Dutch government, the transcripts of the interviews are (in several cases) traceable to specific employees. In order to guarantee anonymity of the interviewees, the transcripts are not provided in the appendix.

### **3.2.3 Methodology**

Analysis of the interviews happened based on the work of Creswell (2016). He suggests different steps that accumulate to a complete analysis of qualitative data. The steps consist of preparing the data, coding the data and interpreting the data. These steps of preparing and coding are considered respectively in the following sections, the steps of interpreting data is covered in Chapter 5.

### **3.2.4 Preparing the data**

The interviews were recorded and during the sessions, short notes were made to keep track of notable passages. Transcribing was done with the help of Microsoft Teams, which transformed recorded audio spoken words of a recording directly into written text. These automatically generated texts were then read and adjusted using recording of the interviews, to correct transcription errors. Subsequently, the interview transcripts were emailed to the interviewees for them to revise their answers and they were asked for consent to use the transcripts in the study. The returned transcripts were used as a basis for further analysis.

### **3.2.5 Coding**

Coding refers to the mechanism allowing to go from raw data to a coherent information. Phrased differently, it is the process of making sense of transcribed data (Chametzky, 2016; Creswell, 2016). Creswell (2016) argues that the process of qualitative data analysis is inductive, which means one starts out with raw data to distill codes from. Constructing codes does thus not happen a priori, but comprises a part of the analysis itself (Weston et al., 2001). It is acknowledged however, that by first having conducted the interviews and later correcting the transcriptions, there already is a premonition for repeating topics and overarching themes in the data.

The process of coding consists of three levels of coding, i.e. open coding, axial coding and selective coding respectively. The first level of coding is open coding, which aims to identify concepts and themes in the data. By reading and re-reading through the transcripts line by line and going through the interview notes, eventually 92 in vivo open codes were created. An overview of the open codes can be found in Appendix II. In the second level of coding,

axial codes were created. These codes aim to ‘further refine, align and categorize the themes’ and this way examine the relationships between emerging themes. The third level of coding is selective coding, in which the axial codes are used to move toward constructing meaning at a more abstract level (Williams & Moser, 2019). Based on the selective codes, a narrative that ties emerging themes together was constructed. Even though this coding process is arguably described quite linearly - moving from one step to the next - reality holds more of an iterative or cyclical process between the steps. This means that in order to get to the final themes and interpretation of the data, the codes and themes were revisited and readjusted multiple times. The axial and selective codes emerging in the data are illustrated in the Appendix III and are discussed in the results section in Chapter 5.

## 4 Results from quantitative analysis

This chapter discusses the results of the quantitative analysis. The results first cover the final sample for the analysis and how this sample relates to complete set of 108 projects. It examines the different predictors and examines whether or not they play a part in predicting the adoption of behavioral insights in policy. The predictors are first discussed individually and later collectively modelled.

### 4.1 Representativeness of the sample

As previously elaborated (par 3.1.2), the data were scored on different characteristics. On the features derived from the BIN NL database - that is effect size significance, use of control group, type of organization, type of intervention and pre-existing communication - the projects in the final sample were compared to the projects not included in the final sample. This way, it was established if there were statistical differences between the initial sample and the final sample and thus whether the final sample represented the total group of projects well.

This comparison was done using a Chi-squared test of independence for each predictor. The results are shown in Table 3. The output shows that none of the predictors is statistically significant, indicating that there are no differences between the distributions of features within the total set of projects and the final research sample.

Predictor	Chi <sup>2</sup>	P
Effectsize significance	.29	.59
Use of controlgroup	1.95	0.16
Type of organization	5.38	0.07
Type of intervention	.49	.48
Pre-existing communication	.04	.82

TABLE 3: TESTING FOR DIFFERENCES BETWEEN TOTAL SET OF PROJECTS AND FINAL SAMPLE

### 4.2 Average adoption

First considered is the average rate of adoption of the projects. Figure 2 presents the different outcome possibilities for the trials with regards to adoption. Overall, 50% (23 out of 46) of the trials was *to some extent* adopted. The distinguishment in the degree to which the outcomes were eventually adopted was described in section 3.1.3.

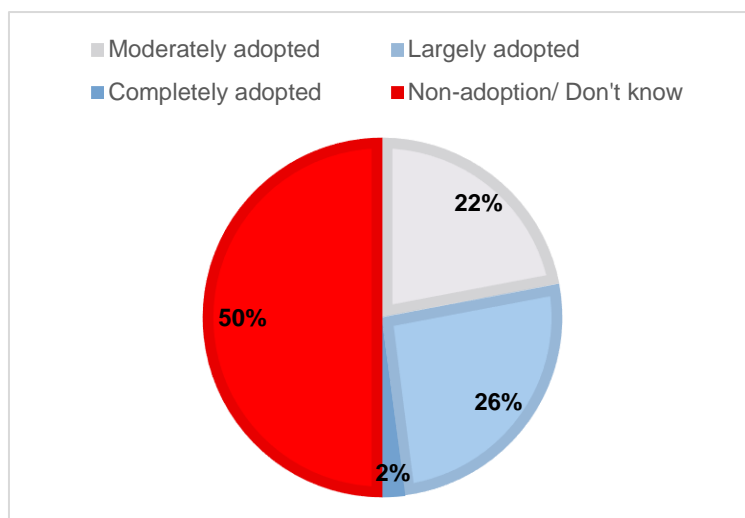


FIGURE 2: OVERVIEW RATE OF ADOPTION

Nine respondents indicated that they did not know if and how the findings of projects were eventually adopted. These responses were labeled as not adopted, meaning that the real percentage of uptake could in reality be slightly higher than what the following numbers suggest.

76% of the trials (35 out of 46) was initially aimed at contributing to policy. As elaborated in section 3.1.4, a two-proportion Z-test was used to calculate differences between groups. From the projects aimed at adoption, 54% (19 out of 35) of the interventions was adopted. From the projects that were not aimed at adoption, 36% (4 out of 11) of the interventions was adopted. Figure 3a indicates that there is no significant difference between the percentage of adoption in the group of projects that was aimed at adoption and the group of projects that was not specifically aimed at adoption.

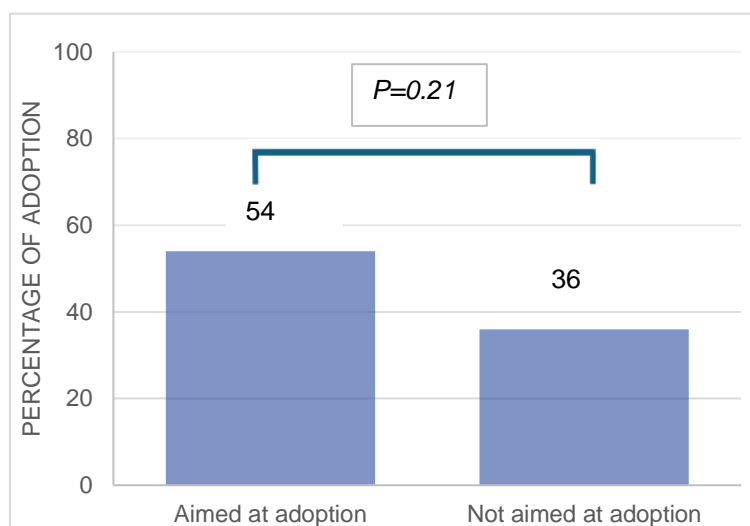


FIGURE 3A: NO SIG. DIFFERENCE BETWEEN AIMED/NOT AIMED AT ADOPTION

The next figure, that is Figure 3b, covers the observed rate of adoption of projects compared to the expected rate of adoption. The observed rate of adoption covers the adoption of all the projects in the sample and the expected rate of adoption is the mean of the expected adoption percentage as given by the respondents in the survey. The first column depicts the mean of the uptake as predicted by the respondents in the survey, which is 39%. The second column shows the actual uptake of the projects in policy, which is 50% (23 out of the 46 projects). The 50% observed adoption does not significantly differ from the expected adoption of 39%.

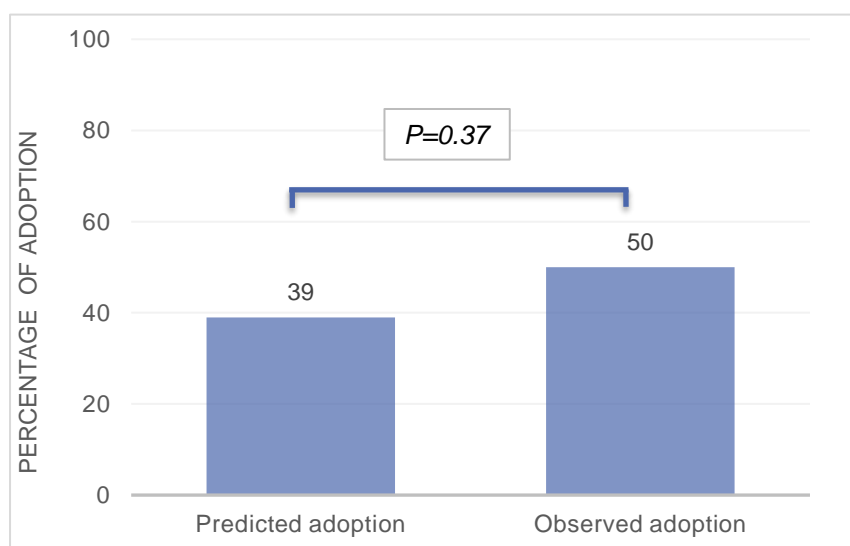


FIGURE 3B: NO SIG. DIFFERENCE BETWEEN PREDICTED ADOPTION AND OBSERVED ADOPTION

### 4.3 Predictor: evidence

The first considered predictor of adoption is the statistical evidence found in the projects. Some of the trials considered more than one intervention and examined which was (most) effective. In these cases, if one of the nudge arms tested had a significant and positive effect, the project is labeled significant. 83% of the projects (38 out of 46) had a significant positive effect. From this group of significant positive effect size, 53% (20 out of 38) of the interventions was adopted. From the projects with no significant positive effect size, 38% (3 out of 8) of the interventions was adopted. Figure 4a shows that there is no significant difference between adoption for these two groups.

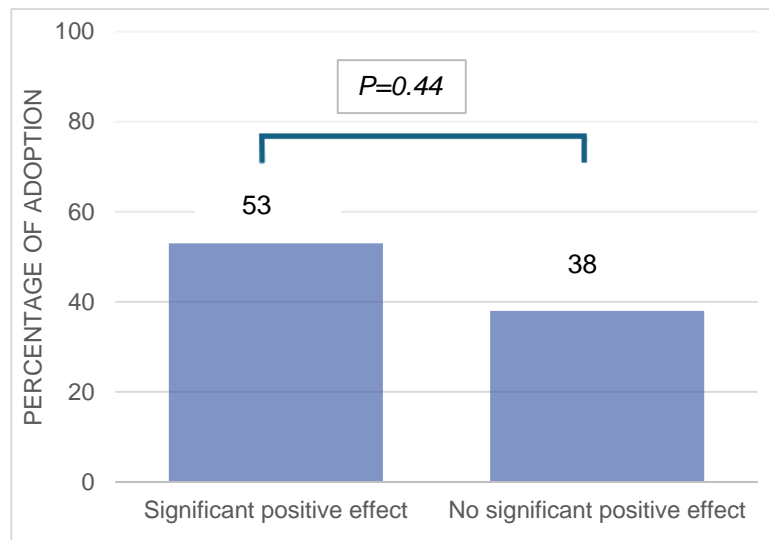


FIGURE 4A: NO SIG. DIFFERENCE BETWEEN EFFECT SIZE

In 83% of the projects (38 out of the 46) a control group was used. Of the projects that used a control group, the intervention was adopted 47% of the cases (18 out of 38), whereas adoption is 63% (5 out of 8) in the trials where no control group used. There is no significant difference between these percentages of adoption as figure 4b presents.

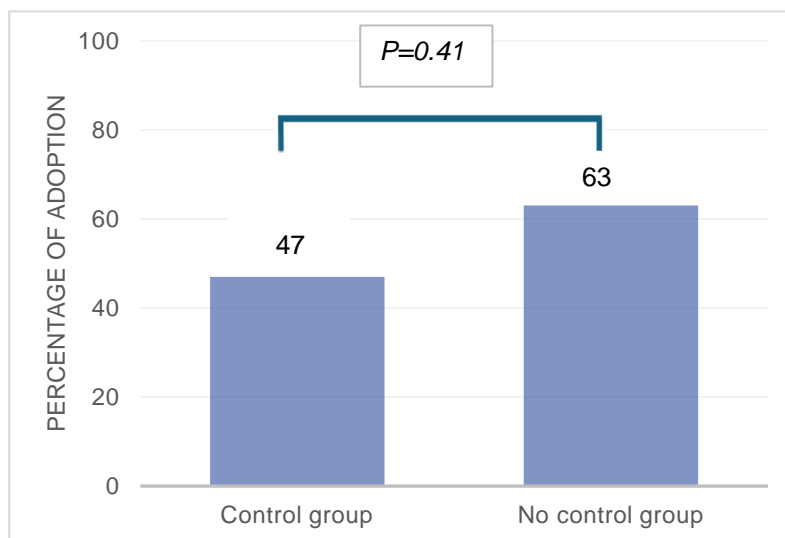


FIGURE 4B: NO SIG. DIFFERENCE BETWEEN USE OF CONTROL GROUP/ NO CONTROL GROUP

#### 4.4 Predictor: type of organization

All BIN NL projects are carried out by a type of (governmental) organization which are distinguished into ministries, executive agencies, regulatory agencies, provinces and municipalities. Following the arguments of DellaVigna et al. (2022), the size of the organization involved could be related to adoption. The organizations participating in the BIN

NL projects differ per project and in several cases a combination of the aforementioned parties is involved. The combination of ministry and other type of governmental organization occurs in 28% (13 out of 46) projects. In this category, 62% of the projects resulted in adoption whereas adoption is 50% (6 out of 12) when a ministry only was involved and 43% (9 out of 21) when another type of governmental organization only was involved. Figure 5 indicates that there is no significant difference between the percentage of adoption of the two most divergent categories.

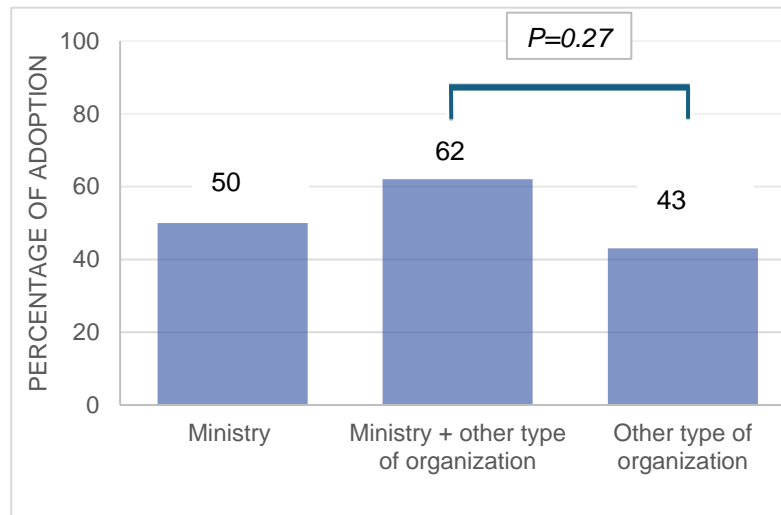


FIGURE 5: NO SIG. DIFFERENCE BETWEEN THE TYPES OF ORGANIZATION

#### 4.5 Predictor: type of intervention

The third category of predictors goes into the type of intervention used in the project. The type of intervention was further classified based on the work of Congdon et al. (2011) and Madrian (2014). A more detailed description of this categorization can be found in Appendix V. This results in three categories of interventions, respectively imperfect optimization, bounded self-control and non-standard preferences. These categories are then split into two groups, based on how invasive the intervention is regarded. Imperfect optimization is linked to interventions considered less controversial, whereas the bounded self-control and non-standard preferences interventions are regarded more invasive (DellaVigna et al., 2022). In some projects there were multiple behavioral interventions used. In these cases, the project was assigned to the category associated with the intervention of the largest impact. Whether or not the controversial nature of an intervention affects the uptake will be covered in section 4.6.

52% (24 out of 46) of the interventions from the sample addresses imperfect optimization, 18% (8 out of 46) addresses bounded self-control and 30% of the interventions addresses non-standard preferences. Of the interventions addressing imperfect optimization 63% was adopted, whereas for the interventions concerning bounded self-control and imperfect optimization respectively 38% and 36% was implemented. Figure 6a illustrates that there is no significant difference between the most divergent percentages within this category.

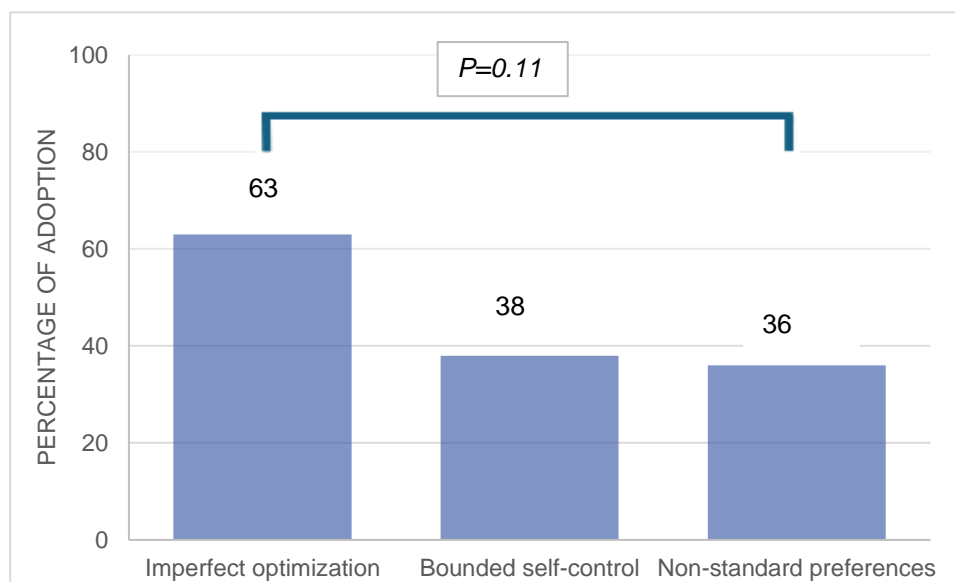


FIGURE 6A: NO SIG. DIFFERENCE BETWEEN TYPE OF INTERVENTIONS USED

In addition, the prevalence of pre-existing communication of the intervention was analysed. Out of the 46 trials, 15 projects had a pre-existing form of communication. The adoption is 40% (6 out of 15) within the category of pre-existing communication. In the projects where there was no pre-occurring type of communication, the adoption was 55% (17 out of 31). Again, no significant difference between adoption between these two groups was found (Figure 6b).

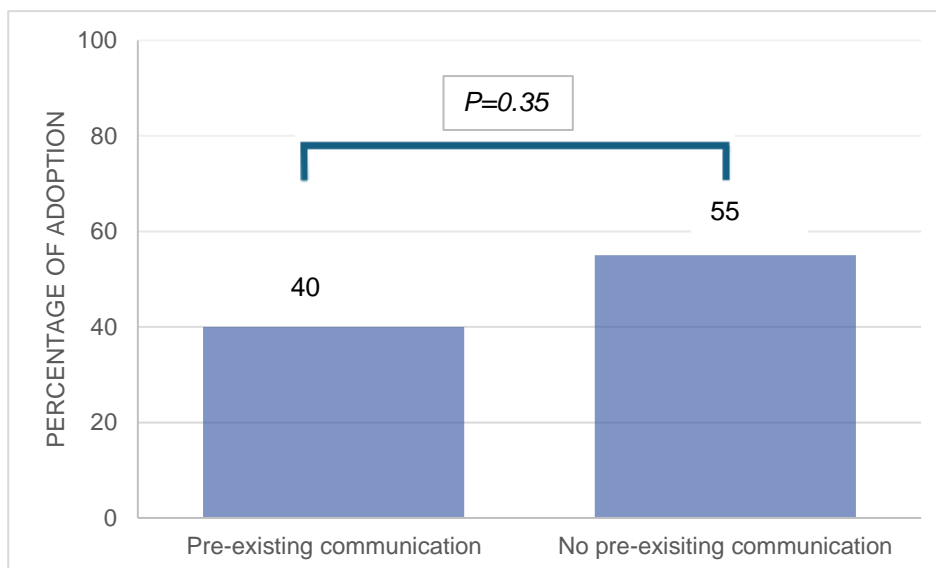


FIGURE 6B: NO SIG. DIFFERENCE BETWEEN PRE-EXISTING COMMUNICATION OR NOT



#### 4.6 Predictors in a logistic regression model

This section covers the logistic regression model, verifying if the independent variables can predict adoption of behavioral insights. The robustness checks for this model are covered in Appendix IV. Table 4 presents the output for the different categories of predictors. For the variable of *initiator of the project*, behavioral advisor was used as the reference category in the model. For the predictor of *organization type*, ministry involvement was used as the reference category.

While interpreting the results from the analysis, the final sample size should be considered. Given that there are 46 observations in total, it is not possible to use a model that captures all the predictors simultaneously, because of the issue of overfitting (Field, 2018). Rather than putting all the predictors together in one model, the predictors are considered in a model one by one, i.e. isolated from the other variables.

Dep. var: adoption (yes/no)	Logistic	Wald stat.	Chi-Square (Omnibus)	95% CI for Odds ratio			Nagelkerke R Square
				Lower	Exp. (B)	Upper	
Aim of project	.731 (.713)	1.053	1.086	-.736	2.078	2.590	.031
<i>Constant</i>	-.560 (.627)						
Effectsize significant	.616 (.799)	.594	.611	-1.152	1.852	21.303	.018
<i>Constant</i>	-.511 (.730)						
Use of control group	-.616 (.799)	.594	.611	-21.357	.540	1.099	.018
<i>Constant</i>	.511 (.730)						
Organization type (1)	.288 (.811)	.157	1.129	-1.463	1.333	1.947	.032
(2)	.758 (.721)	1.105	1.129	-.619	2.133	2.565	.032
<i>Constant</i>	-.288 (.441)						
Intervention type	1.070 (.612)	3.062*	3.173*	-.49	2.917	2.511	.089
<i>Constant</i>	-.560 (.443)						
Pre-existing communication	-.600 (.639)	.881	.895	-2.043	.549	.693	.026
<i>Constant</i>	.194 (.361)						
Initiator of project (1)	2.185 (.880)	6.170**	7.920**	.722	8.889	22.068	.211
(2)	.352 (.790)	.199	7.920**	-2.077	1.422	2.257	.211
<i>Constant</i>	-.575						
Number of projects	46	46	46		46		46

\*=P<.1

\*\*=P<.05

*The standard error is given between brackets.*

TABLE 4: DETERMINANTS OF ADOPTION

The first thing to consider is the overall model significance. This can be derived from the Chi-Square statistic (Field, 2018). Column 3 indicates that the predictor *initiator of a project* can be regarded a significant fit to the data. The Omnibus test compares the used model to a model without predictors beside the intercept, so if this test is significant, the models significantly differ from the null-model, that is a model without predictors. The predictor *initiating the project* shows a Chi-Square value below the threshold of .05, meaning the model significantly improves when adding this predictor. The predictor *Intervention type* yields a Chi-Square value that is below the threshold of .1. This indicates that it is approaching significance and there could possibly be a trend. The larger part of the models does not significantly improve when the predictor variable is added.

The model fit, depicted by the pseudo R Square, can be interpreted by the output values of the Nagelkerke R Square. Considering all the models, the Nagelkerke R Square value for the predictor *initiator of the project* stands out from the rest. The value of .211 indicates that in this model, 21.1% of the variance in the dependent variable (adoption) is explained by the predictor. Generally speaking, this is not strikingly high, but it can be concluded that this variable holds more explanatory value than the other predictors in the model. The explanatory power of the models besides *initiator of the project* can be regarded as fairly low.

The overall model significance (the Chi-squared, or omnibus test) and the Wald statistic in a model with only one predictor, arguably narrow down to the same thing. The Wald test examines whether a specific predictor makes a significant contribution to the prediction of the outcome variable (Field, 2018). In this case, when a single-predictor model is considered one by one, these two tests essentially both examine whether the predictor is significantly useful in the model. When looking at the output values of the Wald statistic, the variable *initiator of the project* is the only significant predictor of adoption. This would indicate that if a project is initiated by a policy officer, it is 8.8 times more likely to get adopted compared to when it is initiated by a behavioral advisor. The bootstrap confidence interval (CI) is added to the model as an extra robustness check. Regardless of the broad CI range (indicating that the population value of  $b$  falls between .722 and 22.068 95% of the time) corresponding with *initiator of the project*, the interval does not include zero. This means that the positive relationship between *initiator of the project* and the outcome variable is confirmed. The Wald statistic for *intervention type* is not significantly contributing, but it is below the threshold of .1 and thus approaching significance. There could therefore be a trend between the use of a non-controversial intervention type and adoption.

A final remark about the output of the logistic regression concerns the negative signs before the *b* of *use of control group* and *pre-existing communication*. This negative sign is reflected by the exp. (B) values which are below 1 and implies a negative relationship between these variables and adoption, which is not in line with the hypotheses.

## 4.7 Open-answer questions

In the survey, respondents were asked what they believed to be the main reason for non-adoption for interventions from the projects. There are different explanations, which are discussed underneath.

The first explanation is that actually implementing and adopting the findings of a project proves to be more difficult than ‘just’ performing a pilot. This is partly due to problems with scalability and financial resources. It is also stated that recommendations that are easy to implement are more likely to be adopted compared to insights that require a systematical change, because it is hard to stay involved and find support for these things.

In addition to that, the environment in which policy is developed can be a hindering factor. When developing policy, there are many stakeholders and thus many interests involved, and the behavioral component is not always on top of the priority list.

It is also noted that the political trust in traditional policy instruments is larger than in behavioral instruments. Consequently, policy professionals tend to prefer traditional instruments over the use of behavioral insights, even if the effectiveness of these instruments is doubted. This complements the prior explanation, in the sense that the insights are not valued to the extent that they will be actually used.

### 4.7.1 AI verification

As explained in Chapter 3, the open answers as filled in in Qualtrics were entered into ChatGPT as a form of explorative validation of the initial interpretation. The main themes according to ChatGPT are:

- Trust in traditional solutions – Policymakers and politicians prefer traditional government solutions such as information dissemination, knowledge transfer, regulation, enforcement and subsidies. There is a clear preference for certainty, even when the effectiveness of these traditional methods is questionable. This stems from a lack of trust in unfamiliar and experimental behavioral approaches.

- Outcome and implementation issues – Although some projects yield positive outcomes, these results are often not scaled up or correctly implemented. This is largely due to personnel changes and a lack of basic knowledge about behavioral change among internal stakeholders. These factors lead to a loss of collective memory and orphaned ownership, jeopardizing the continuity and success of integrating behavioral insights into policy.
- Practical barriers and political will – significant practical and political obstacles hinder the integration of behavioral insights. Time constraints, political context, and the timing of research within the policy cycle play crucial roles. Moreover, implementing behavioral insights often requires fundamental changes that are not always feasible within existing policy structures and political will.

The following chapter further examines the explanations behind (non) adoption of behavioral insights in policy.

## 5 Results from qualitative analysis

This chapter covers the findings of the analysis of the interviews. The interviews examined both bottlenecks and success factors that either interfere with, or contribute to adopting behavioral insights in policy. The representativeness of BIN NL projects is considered first. Secondly, several general barriers of implementing behavioral insights will be discussed. Furthermore, the chapter considers experienced bottlenecks concerning implementing and holding on to initiatives from projects that are deemed effective. In addition to that, factors that facilitate a smooth uptake of behavioral insights according to behavioral experts are debated.

### 5.1 Representativeness

In order to verify the representativeness of the BIN NL database for all behavioral research, the interviewees that also contributed to BIN NL research were asked to what extent they deemed projects published by BIN NL representative for behavioral research in general. 9 of the 12 interviewed behavioral advisors had contributed to a project published in the BIN NL database. It is stressed by several advisors that the BIN NL projects do not encompass all work that is done regarding behavioral insights. First of all because these projects are all some form of an experiment, meaning that (most) campaigns, strategic documents or preliminary findings are by definition not included. The BIN NL database can thus be regarded as a simplification of all research. In addition to this, some projects are withheld from publication by the originator, because of topic sensitivity [R3, R5, R6, R10 & R12].

While acknowledging the caution to describe the BIN NL database as a completely comprehensive representation, it is mentioned that the BIN NL published projects for the larger part do represent the research conducted well [R2, R3, R4, R6, R7, R11 & R12]. “*Almost every project we do, we publish there* (referring to BIN NL)” – R2. Indeed, to some degree, the aim of doing research is to get it published by BIN NL [R4]. The researches as published by BIN NL are considered the most concrete examples and the cases that are actually followed up on [R11 & R12]. With regards to the impact of the projects, it is mentioned that the impact as published might be smaller than the actual impact, considering that impact in the form of motivating coworkers to use behavioral insights in their work is not incorporated as it is hard to distill [R2].

## 5.2 General barriers of adoption

There are several studied barriers emerging with the adoption of behavioural insights in policy. The extent to which these barriers occur are presented in the following sections.

### Lack of time

Lack of time as a barrier is twofold. On the one hand there is a mismatch between the timeframe of policy development and the timeframe of conducting research to provide evidence for an effective intervention. Often when behavioral advisors are consulted, a quick answer or solution to the issue is expected. These expectations stem largely from political pressure but cannot always be met [R2, R4, R8, R9, R10]. A second aspect of the barrier of time is the time employees have available to dedicate to a certain project. Because of other work-related commitments and tasks of the policy professional, time for examining behavioral aspects of an issue is often not available [R5, R7, R14]. Lack of time as a barrier is validated by all respondents.

### Lack of expertise

The policy professionals working to develop policies are only to a restricted extent familiar with working with behavioral insights. Unfamiliarity with the application of behavioral insights traces back to different actors involved, not only to the ones concerned with policy development, but also with the ones in charge of making final decisions and the ones responsible for implementation and execution. A lack of basic knowledge and a lack of behavioral experts involved in this process complicate the adoption of behavioral insights in policy [R1, R2, R3, R7, R9, R10, R11, R15]. In addition to that, the amount of behavioral experts working in the organization is limited, meaning that the BITs are relatively a very small part of the organization. BITs are hence not able to support every part of the organization or help out all employees [R3, R4, R5, R8, R14].

### Lack of reward & lack of stimuli from management

It is recognised that management does not actively support the use of behavioural insights [R3, R7, R8]. In this sense, the bottleneck appears not to be on the level of the policy officers, but rather on the management level that is above. *“Often, it is the policy officers who have to urge their supervisors to allow them to proceed”* – R8. The managers or department heads

themselves are not always keen on the application of behavioural insights, so they arguably do not motivate their policy officers to do so. *“Policy officers are usually willing, but they have their department heads above them. (...) Behavioral approaches are usually not well received, because management likes to play it safe”* – R7. Indeed, in some cases it is negatively valued if employees include the behavioural component in their work. *“You’re better off avoiding it (referring to behavioral components), as then you do a better job compared to when you do take them into account”* -R14. The behavioural component is in these cases regarded as complex and time consuming and hence no favourable option for a policy response.

#### Lack of autonomy

The lack of possibilities for policy officers to make their own decisions was not confirmed specifically by the interviewees. It was mentioned that policy officers find it important to make deliberate choices, but in practice they are hesitant to do so [R7]. It is acknowledged that people tend to make choices based on how ‘it was always done’. This automatic kind of decision making might cause individuals to feel like they do not have autonomy to make decisions, even if in fact they do. In addition to that, a certain degree of reactance among employees (medior and senior positions specifically) was acknowledged.

#### Lack of embedding in the workflow

The lack of using behavioral insights embedded in the workflow is largely acknowledged by interviewees [R1, R2, R3, R7, R8, R13, R14]. At the same time, a there appears to be an increase in degree of embeddedness of behavioral insights in the policy development process. The introduction of the “policy compass” (beleidskompas), a relatively new guidance tool developed in the Dutch national government that policy officers use when developing policies, contributes to this shift. The effect of the policy compass on actual application of behavioral insights should however be nuanced. On the one hand, it aims to make policy officers think about the behavioral components of a certain policy. At the same time, it is mentioned that the introduction of the policy compass as a tool for using behavioral insights is not enough [R2, R3, R10, R15]. Not all employees use the compass, and even if it is formally used (i.e. ‘the checking of a box’) it is not always deliberately used (i.e. the behavioral component is actually taken into account in the policy). If not used deliberately, the format contributes to a more cumbersome and bureaucratic work process rather than the exploitation of behavioral insights.

### 5.2.1 Additional barriers

In addition to the previously mentioned barriers, respondents were asked what other difficulties they encountered with regards to implementing behavioural insights in their field of work. These additional barriers go into the current political preferences and institutional features of organizations. They are presented below.

#### Political barriers

The adoption of behavioral insights depends heavily on the political climate and the preferences of the one in charge of policy decisions [R1, R3, R7, R8, R11, R13, R14]. If in this position, the importance of using behavioral insights is recognized, this positively contributes to the use of behavioral insights and vice versa. In cases where the importance of using behavioral insights is not acknowledged by the executive position, employees struggle to get behavioral insights adopted. *“We’re already preparing for a government that is scared of the word behavioral influence and the effect that will have on all BITs. That our researches won’t be shared proudly anymore”* -R3. The size of the perceived struggle varies, going from experiencing a lack of support to a strong resistance against using behavioral insights.

#### Process & institutional barriers

Another difficulty comes in the shape of different ‘languages’ policy professionals and behavioral experts speak [R6, R9, R11 & R12]. Policy professionals are in their work in general more pre-occupied with political sensitivity, whereas behavioral experts stress the importance of evidence-based working. This indicates a communication gap: a clear connection between these fields is still lacking, complicates engaging with an evidence-based approach for policy officers.

When aiming to implement behavioral insights in policies, it arguably helps if the policy professionals are able to find the BITs or behavioral experts in their organization. Right now, this happens for a large part based on (positively) previous relationships. *“Some departments or colleagues have experience with us that turned out well. This is important, as they will find us again”* – R8. The explanations for difficulties regarding finding the right person (i.e. the behavioral expert) is firstly because employees within the organization are simply not aware of their existence. Secondly, even if employees are aware of the behavioral



insights knowledge (either in the shape of teams or individuals) within the organization, it is not always clear to them how to find the right people and get in touch with them.

### **5.3 Barriers of adoption – in case of effective interventions**

If an intervention is proven effective in the experimental setting, then what prevents this arguably successful intervention from being adopted in policy? Several explanations come up when looking at the data.

#### Organizational barriers

Before an intervention is actually implemented, there are various people engaged in the process of granting the ‘final permission’. This involvement of a variety of people leads to the crowding out of the initial behavioral intervention [R1, R3, R15] . *“Often there are a lot of different people involved, who are kind of related to decision making. (...) A lot of consents are required and a lot of people give their input, which basically always causes something in the process that crowds out the behavioral aspect.”* - R1. This is closely linked to the hierarchy in organizations. If the person with executive power does not favor the intervention over another kind of policy, either because they do not believe it to be effective enough or because they are not willing to spend the financial resources, it will eventually not get implemented. The more administrative hoops an intervention has to go through, the more likely it becomes that at some point, someone is going to disagree with the intervention or alter it.

Another aspect of the organizational barrier is the compatibility of the behavioral insight in the organization. If the innovation does not ‘fit in’ right away, it is more likely that it gets stuck in the system, and this way not make it to implementation [R4, R5].

Moreover, the success of implementing an intervention in many cases hinges on the ‘knowing the right people’ [R1, R2, R8, R9, R10, R11, R14]. Behavioral experts have to know the right people, in other words a place for the intervention to land, to realize a follow up on their tested intervention. Respondents indicate that in the process of bringing a trial to actual policy, they depend on a (previously built) relationship with policy professionals. When policy colleagues have good experiences with the behavioral expert, they are more willing to engage.

#### Political barriers

The political barrier experienced is in line with the barrier previously described. If policy officers want their policy idea to be approved – whether this is on the level of the central

government or local government – it must undergo some form of verification by politicians. If the general rationale is that politicians are not keen on implementing behavioral insights, or that these ideas are not on top of the priority list, policy officers shy away from proposing behavioral intervention – whether they are proven effective or not [R3, R7, R8, R9, R11]. *“When it comes to politics, you’ll have to tell them what they want to hear. Otherwise it (referring to an intervention) will just get shut down”* - R7. This builds upon the condition that individuals in power positions have to acknowledge the importance of behavioral insights before they can actually be adopted. *“Some people are just not that fond of research, but they do hold a power position where they are in charge of certain decisions. And if they are just not that much of a fan, it’s possible it (referring to the intervention) will not be used”* – R11.

### Scalability barriers

Interventions are usually tested in the environment of the project, for example in a pilot. Issues emerge when extrapolating the project from a pilot setting to larger scale. Due to a lack of recourses, usually in the form of financial funding and expertise required throughout the scaling process, innovations do not always make it to a larger scale. *“If you want to scale to national level, it turns out there are no funds available. (...) Then they’ll tell you it’s not going to work and the behavioral project grinds to a halt”*- R7.

Whenever an intervention is deemed effective, this effectiveness is observed in the context of a trial. The context dependency of the intervention is an argument for hesitant employees not to engage with the use of behavioral insights. Regardless of the proven effectiveness, the policy setting the intervention is meant for is arguably different from the experimental setting, causing doubts concerning the effectiveness in practice [R6, R10, R15].

#### **5.3.1 Longevity of interventions**

Respondents were asked what difficulties they experience when it comes to making an intervention last. Two barriers were identified.

### Turnover in the workplace

The main factor that negatively affects the longevity of an implemented intervention is turnover in the workplace. *“Sometimes you notice a new person on the case. (...) So the results you tested were implemented, but the knowledge gets lost in the process anyway, because a new person will use a different approach”*-R2. When an innovation is successfully implemented and the

involved employee leaves, this causes a disruption. They often take with them valuable knowledge about the policy idea and this knowledge concerns not only details about the implementation, but also the underlying ideas of the intervention [R2,R4, R5, R7, R13, R15].

### Development over time

An additional explanation for the possible fading of an intervention in policy is the passing of time. As time passes, the intervention does not stay on top of mind by employees and attention slowly dims. New developments come up and consequently, the intervention becomes less relevant. [R2, R4, R5, R6, R9, R15].

## **5.4 Success factors**

There are four success factors distilled that help contribute to a successful implementation of behavioral insights. These success factors are presented in this section.

### Commitment in the organisation

Another important success factor is the commitment within the organization. This commitment is highly relevant on both on the level of the policy officers and on the management level. Managers committed to work with findings of the projects and having this commitment established beforehand is an important factor that contributes to adoption. *“When there is commitment - especially on the level of higher management - to take up the projects, the chances of success are much higher”* – R7. To prevent the intervention from ending in a void and thus not being utilised, the importance of finding a so called ‘place to land’, preferably in the shape of buy-in on management level, is emphasised.

Besides the emphasis on the importance of cooperation on management level, the policy officers are of great relevance. *“The enthusiasm and commitment of the policy-coworkers is really important. They are basically the gate-keeper of how well things develop”* - R2. In their role, both the willingness to cooperate and a basic level of behavioural knowledge is considered of use to help implement behavioural insights. This can be explained by their position within the organisation, in which they develop and shape policies.

Along the lines of commitment, the importance of ambassadorship and stakeholder management is stressed as well [R1, R4, R5, R10, R11]. This ambassador can ‘sell’ or advocate for the interventions to people in the organisation who are in these power-positions. Ideally however, the ambassadors have the authority to make decisions themselves. The ambassador

can take for example the form of a manager or a department head who is both enthusiastic and has some kind of intuition what can be expected from the project. “*You always need some kind of ambassador who’s in your favor – and who can decide*”- R8. In addition to this, the importance of stakeholder management is emphasized [R2, R4, R5, R6, R7, R10]. Taking into account early on which stakeholders are relevant for the follow-up process, pinpointing the important parties and guiding them through the entire process can yield major benefits at the end of the trial in terms of adoption.

### Sharing good results

The respondents deem presenting the effects as tangible or concrete as possible a success factor for the implementation of an intervention. Tangible results can be used to prove the effectiveness of using behavioral interventions and this way convince hesitant people within the organization. In addition to that, it is emphasized that the good results should be shared and success stories should be told, in order to help the adoption of behavioral insights gain traction.

## 6 Discussion and conclusion

This chapter provides an answer to the research questions “*To what extent have findings of behavioral insights projects been integrated in policy initiatives in the Netherlands?*” and “*What bottlenecks do policymakers face when adopting behavioral insights in policy?*”. The conclusions are related to other research findings in order to place them in perspective and the key strengths and limitations of the study are identified. The chapter concludes with a policy recommendation and further research recommendations.

### 6.1 Conclusion

This study examines the uptake of behavioral insights in policy in the Netherlands and the barriers policymakers face when implementing behavioral insights. The quantitative leg of the study examined the behavioral insights database as composed by BIN NL, containing behavioral studies conducted throughout the years 2014 - 2024. In addition to the quantitative set up, a qualitative approach in the shape of interviews was used to shed a light on the bottlenecks and success factors policy officers face regarding the adoption of behavioral insights.

#### 6.1.1 Extent of policy adoption

The findings of behavioral projects in the Netherlands between 2014 and 2024 are to a moderate extent adopted in policies. In 50% of the cases, the results of a project are *to some extent* used in a follow up policy initiative. There were no significant differences found between the total set of projects and the final sample used, regarding the project features. No significant effects of the predictor variables *significant effect size*, *use of control group*, *organization type*, *intervention type* and *pre-existing communication* on the adoption of behavioral insights in policy were found. *Initiator of the project* was the only independent variable that was found to significantly contribute to predicting the outcome variable. To be more specific, if the initiator of a project is someone in a policy-related position (i.e. someone concerned with policy implementation, policy supervision or a policy advisor), the likelihood of adoption increases. The predictor of *intervention type* was found to approach significance, suggesting that there may be a trend or an effect going on.

### **6.1.2 Explaining policy adoption**

There are different barriers emerging when it comes to the application of behavioral insights in policy. The most recognized general barriers were a lack of time, lack of expertise and a lack of embeddedness of using behavioral insights in the work processes. Added to these barriers were political and process or institutional barriers. The political barrier captures the fact that adoption of behavioral insights for a large part depends on political preferences or the preferences of individuals in power positions. The latter type of barrier refers both to the discrepancy between languages policy officers and behavioral advisors speak and the ability of policy officers to consult behavioral experts.

When it comes to adopting findings of projects that are considered effective in the experimental setting, there are two barriers mentioned specifically that are in line with the previous two barriers. The first barrier captures the organizational issues, referring to the crowding out of initial behavioral interventions due to noise that comes hand in hand with including many different individuals with the eventual decision. Another aspect of this organizational barrier is the issue of knowing the right people in order for an intervention to land and the suitability of the intervention in the existing system. The second barrier covers the political context in which policies are developed and is in line with what was previously described. Policy officers are inclined to tell politicians (or others in charge) what they want to hear in order to get their idea approved, regardless of the established effectiveness of an intervention. Barriers addressing the longevity of interventions occur in the shape of turnover in the workplace and new innovations emerging with developments of time, which makes it hard to hold on to implemented interventions in the long run.

There are two success factors identified that contribute to an arguably improved adoption of interventions. The first one concerns commitment within the organization. This commitment is preferably established on management level, or other people who are in charge of final decisions. The main idea is that commitment on higher levels within the organization can secure a place to land for the intervention. The willingness, the engagement and a sufficient base of behavioral expertise of policy officers also largely contribute to a successful implementation. The final factor is about the promotion of good results: by sharing success stories and showcasing effective results, the adoption of behavioral insights can gain traction.

It can be concluded that there are several factors complicating the adoption of behavioral insights in policy. These factors do not exist independently, but are to a large extent interconnected. When examining tested interventions specifically, the main hindering factors

that prevent interventions from being followed up by policies, are organizational features and the political environment of policy development. Turnover in the workplace and new developments are the main reasons for not being able to hold on to implemented interventions in the long run.

## 6.2 Discussion

### Quantitative results

The sample size of this study should be kept in mind when interpreting the results. As elaborated, a smaller sample size comes together with decreased statistical power, meaning it is harder to detect effect sizes (Haile, 2023; Wilson-VanVoorhis & Morgan, 2007). Apart from one exception, the used predictors did not improve the overall logistic regression model compared to using no predictors. The dataset could however just be too small to prove this (possibly existing) effect. The same applies however for the results of the calculated differences between the final sample size and the initial set of projects (Par 4.1). Again, no effect was observed, meaning that on first glance it appears as if the sample represents the total dataset well, but this dataset too could be too small to prove an effect, if it exists. Another explanation for the lack of significant results could be that the predictors in the model do not actually comprehend the underlying mechanisms. This would imply that the dimensions that in fact explain policy adoption, are not captured by the variables.

One variable used in the logistic regression model was found significantly improve the used model, that is the predictor *initiator of the project*. The predictor of *intervention type* was found to have approaching marginal values in the Wald test, implying that there may be a relationship. but considering the final sample size, When comparing these findings to the study of DellaVigna et al. (2022), there are both similarities and differences. First of all, the results of DellaVigna et al. (2022) were also for the larger part not significant. Contradicting this study, the strongest evidence they found was for the predictor of pre-existing communication which was followed by the strength of evidence. Based on these findings they formulated the idea of organizational inertia as a barrier for the adoption of behavioral insights. The findings of this study do not validate their results.

The significant value of the predictor *initiator of the project* makes sense, at least in theory. The output showed that if a project is initiated by a policy officer, the chances of adoption in policy are higher compared to when a behavioral advisor initiates the project. This could be explained by the fact that the policy officer is in a better position to actually use the

findings. In addition to that, if the policy officer initiated the project, it is probably safe to assume that they are in fact interested in the results and how to apply them. Again, given the limitations of the dataset, these results should be interpreted with caution.

### Qualitative results

The two overarching themes emerging regarding adoption of behavioral insights in policy are organizational barriers and political barriers. The most emphasized barriers that followed from the interviews are in line with what could be expected when looking at the literature. For example, the issue of scalability as illustrated by List (2024) is apparent and poses a recognized bottleneck when it comes to adopting behavioral insights. Moreover, the barrier of the political landscape was described by the respondents as well. Some of the identified barriers are more general than others. For example ‘knowing the right people’ is not specifically limited to implanting behavioral insights, but is applicable for other policy initiatives as well.

When taking a step back and looking at these results, there appears to be a hiatus in theoretical importance of behavioral insights - which is broadly recognized, also by the Scientific Counsel for Government Policy (WRR) - and the experienced importance of behavioral insights by behavioral advisors and policy officers. There is willingness, but a lack of knowledge and expertise on the level of policy development, but resistance on management levels and individuals in power positions. This unwillingness of the management level seems rather contradictory at first sight: on the one hand, there is the government-wide implementation of the policy compass tool, which urges policy officers to take into account behavioral components in the policy development process. Yet, when behavioral advisors urge to opt for a certain (behavioral) policy direction, this is in many cases subordinate to other priorities.

To describe the sharing of effective results as a success factor, seems contradictory to a certain degree. Several respondents indicated that regardless of the evidence of interventions, they might still not get adopted because of other present priorities. Indeed, even if it is known that a certain policy might not be effective, it can still get implemented, for example for the purpose of sending a message to public. Considering this, it can be concluded that the effectiveness of an intervention is – in certain cases - subordinate to other interests. Nonetheless, advocating for the application of behavioral insights by showcasing good results, can intuitively help convince people.



### 6.3 Strengths & limitations

One of key strengths of this research is that it is innovative, in the sense that it studies an area about as if yet is little known. Some (quantitative) research into evidence adoption in policy exists, but not specifically focused on the Dutch context. Another strength of this study regards the combination of a quantitative part and a qualitative part. Combining these methods allows for a more comprehensive and in-depth understanding of the topic. There are also several caveats of this study that should be considered.

The first limitation addresses the way of operationalizing the research question. Adoption is measured by means of the uptake from BIN NL projects specifically. Not all behavioral researches however end up in the database used. Indeed, the BIN NL database solely focusses on the experimental set up. The BIN NL database can thus not be considered as a proxy for all behavioral research conducted in the Netherlands thus far. This limitation of operationalization is acknowledged, and the study does therefore not make any generalized claims about all behavioral research. The conclusions are tailored for the specific projects showcased by BIN NL.

To mitigate this limitation, the representativeness of the BIN NL projects was measured in the interviews. This was done by asking the interviewees that submitted projects to BIN NL to what extent they deem these BIN NL published projects and their impact representative for the research they generally participate in. The presented results show that given the specific features the projects have to match to be published, and underrepresentation of behavioral research does not seem apparent. Indeed, an overrepresentation or publication bias would seem more likely, considering the emphasis of respondents on how they were most likely to present their ‘big cases’ or ‘concrete examples’ to BIN NL. The validity of the quantitative study was also secured by the check in differences between the final research sample and the entire dataset of BIN NL (see section 4.1).

On a further note, while recognizing that the BIN NL database might not cover all behavioral research, it is the most broad comprehensive database available when it comes to researching behavioral initiatives. In other words, it is unlikely that there is a comparable amount of research conducted, that is not getting published by BIN NL. There was hence no other dataset available that was considered a better option for this study.

Another limitation concerns the measurement of policy adoption in the quantitative analysis. In order to examine the uptake of findings of the trials, this study relies on self-

reported results from the authors of the projects. Addressing the contact persons as listed by BIN NL was deemed the most efficient way to get in touch with the relevant actors, but these authors are not necessarily involved in (possible) implementation phase of the intervention. As a result, some respondents indicated that they did not know what eventually happened to the project findings. Building upon that, the measurement of the extent of adoption could be perceived as rather subjective. Authors were asked in the questionnaire to what extent the findings were implemented, but the answers are considerable subjective. What one person deems 'to a large extent', could someone else perceive as 'slightly', and vice versa.

Furthermore, the proxy for predictor 'state capacity' in this study is arguably not as comprehensive as the variable was in the paper of DellaVigna et al. (2022). To be more concise, what is missing from the data is an element that captures whether or not the person involved in the project still works in the organization. Even though DellaVigna et al. (2022) could not significantly relate personnel with expertise and institutional memory still working in the same position to a higher rate of adoption, it is a plausible line of reasoning. Considering the interviews, where turnover in the workplace was mentioned regularly as a barrier for adoption, it is not unlikely that this factor in fact related to adoption.

On a final note, it is acknowledged that the assumption regarding independent observations is to some degree violated, as explained in Appendix IV. At the same time, there was little to no room to mitigate this violation. The most applicable possible option would namely be to omit certain observations from the analysis, but given the already limited amount of data it was chosen not to do so. The assumption of multicollinearity was not violated according to the output in Table 5 and Table 6. The caution used before with interpreting data applies here as well: in a larger dataset, there could possibly be more significant (problematic) correlations between variables.

#### **6.4 Research recommendations**

Notwithstanding the answers this study provided, it also yields additional questions. These questions can in turn be converted to further research recommendations. Considering both the results from this study and the study of DellaVigna et al. (2022), additional research would be specifically interesting to form hypotheses regarding behavioral insights and its use.

One suggestion for additional research would be to take the quantitative approach that was used in this study and in DellaVigna et al. (2022), and perform it on a larger scale. As elaborated prior, the limited sample size in combination with the logistic regression model is

likely to explain the lack of significant effects in some parts of the data. The predictors of intervention type and initiating project could however be labeled as marginally significant, and thus could indicate a possible trend. It would be interesting to see if this trend – or maybe more than a trend - is also visible when studying at a different scale or area. Performing this study on a greater scale would not only allow to better infer possible relationships, but also to take into account nuances of the features of the projects - for example the extent to which interventions are implemented - in the analysis. Possible study subjects for this kind of research could be the United Kingdom - one of the pioneers when it comes to the application of behavioral insights in policy - or a collective of European countries.

A second suggestion for further research could be to address the final caveat mentioned in the limitations section. The scope and research aim of the study could remain the same, but taking more time to conduct the research, the predictor whether or not a person is still employed at the organization could be taken into account. Especially given the large role turnover in the workplace plays in the longevity of an intervention once implemented, it would be interesting to examine the effect on adoption.

An additional suggestion for further research could be to map or identify the status of behavioral insights adoption in policy in the Netherlands a broader sense. This study focused on the projects published by BIN NL and it is acknowledged that there could be behavioral insights research not covered by this database. It is however not unimaginable that behavioral insights end up in policy without specific prior research. This could happen for example in the case of one project yielding results applicable for multiple policy situations, or because of interests of policy officers that aim to integrate existing insights. The emphasis of this possible study would hence be on policies itself, rather behavioral insights research. To narrow down the scope from all policies to something more manageable, an idea could be to focus on specific policy areas or specific municipalities.

Furthermore, it would be interesting to go beyond the scope of adoption of behavioral insights in policy, and focus more on the effectiveness. By selecting certain behavioral-based policies and measuring their effects, a possible comparison can be drawn between the traditional instrument and the behavioral instrument. Especially considering the issue of the possible fleeting nature of behavioral insights that is mentioned in Chapter 2, effect studies would be highly relevant. These types of comparisons can support the value of behavioral insights in policies and thus help contribute to a shift towards more behavioral-based policies.

A final suggestion for further research could be to examine the adoption of behavioral insights per policy area rather than all together. This could be done along the lines of the World Health Organization (WHO) status report (WHO, 2023), presenting the use of behavioral and cultural insight activities in public health authorities in Europe over the years 2021 and 2022. Studying the uptake and the specific barriers per sector could potentially lead to different, but more specific results. Gaining insight in these barriers allows to draft tailored tools to address them and this way, enhance the eventual uptake of behavioral insights in policy.

## 6.5 Policy recommendations

Based on the findings of this study, it can be concluded that there is room for improvement when it comes to both the utilization of behavioral insights and a government-wide acknowledged importance of behavioral insights. This results in several policy implications.

The first recommendation addresses the barrier of holding on to implemented interventions. Even though the process of getting to implementation might be the biggest hurdle, it was described that turnover in the workplace is a large factor in explaining the fading of interventions. With the leaving of an involved employee, both knowledge about the intervention and the mechanisms behind it and the feeling of responsibility, are more likely to fade as well. The issue of knowledge leaving the organization along with people is of course not limited to this specific area of expertise. In order to increase the resilience of interventions, it would be profitable to have more attention for securing them, once implemented. To lessen the impact of employees leaving, it is suggested that the work regarding the implementation is taken up in small teams of people. The teams should be small enough to keep a sense of responsibility for each individual, but create a shared ‘ownership’ of the intervention at the same time. In the case of someone leaving, there will still be at least someone left within the organization feeling responsible, meaning that the intervention is secured to some extent.

Furthermore, the gap in intention of applying behavioral insights there appears to be between behavioral advisors and policy officers should be addressed. To put it bluntly, it appears as if on the one hand, there are the behavioral advisors, trying to sell ‘their’ ideas, whereas on the other hand, there are the policy officers and people in position to take up these ideas, showing little to no interest<sup>4</sup>. To move out of this impasse, there are arguably different steps that can be taken. First of all, the behavioral knowledge in policy institutions should be

<sup>4</sup> Acknowledging, of course, that reality is more nuanced than this. But the analogy remains applicable.

increased. Not only in the sense of broadening the behavioral teams, but also integrating a certain base of knowledge in the people in policy positions. Additionally, the data has shown that merely 'having knowledge' on the work floor is not enough, as a large part of the barriers stems from resistance from people in power positions. In the painted scenario, employees shy away from presenting behavioral-centered policies, because they know in advance that they are likely to be rejected. By turning this around, and ensuring that incorporating the behavioral component is not only valued but maybe even required, employees get a different incentive. This goes beyond the idea of the policy compass, as the focus should be more on the validation from people in higher positions. The application of behavioral insights should thus be incorporated both top-down and bottom-up.

A final note with regards to improving the incorporation of behavioral insights is the following. Including behavioral components from the very start of designing policy is more effective than trying to incorporate them in hindsight. To help realize this, a starting point would be to institutionalize the BITs. Currently, employees within the organization struggle to find behavioral advisors, if they are aware of their existence at all. Formalizing these teams within the organization could address this issue and increase the findability of behavioral advisors, which in turn could contribute to a higher degree of integration of behavioral insights in the work processes. Having behavioral insights embedded to a higher degree means it will be easier to be taken into account when the policy is still on the drawing board.

## 7 References

- Ahluwalia, M.S. (2015). Role of economists in policy-making. (Working Paper 2015/144). United Nations University World Institute for Development Economics Research.
- Al-Ubaydli, O., Lee, M.S., List, J.A., Mackevicius, C.L., & Suskind, D. (2020). How can experiments play a bigger role in public policy? Twelve proposals from an economic model of scaling. *Behavioral public policy*, 5(1), 2-49. doi:10.1017/bpp.2020.17
- Alm, J. (2017). Is economics useful for public policy? (Working Paper 05/2017). Victoria University of Wellington.
- Agner, E. (2019). We're all behavioral economists now. *Journal of Economic Methodology*, 26(3), 195-207.
- Ball, S., & Head, B.W. (2021). Behavioral insights teams in practice: nudge missions and methods on trial. *Policy & Politics*, 49(1), 105-120.
- Banuri, S., Dercon, S., & Gauri, V. (2017). *Biased Policy Professionals*. (Policy research working paper. No 8113). World Bank Group.
- Behavioral Insights Network Nederland. (2023). *Rijk aan gedragsinzichten – Editie 2023*. Behavioral Insights Network Nederland: Ministry of Economic Affairs and Climate Policy.
- Behavioral Insights Network Nederland. (n.d.). *Over BIN NL*. Retrieved May 13th, from <https://www.binnl.nl/over+ons/over+bin+nl/default.aspx>
- 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>
- Benjamin, D.J., Berger, J.O., & Johannesson, M. *et al.* Redefine statistical significance. *Nature Human Behavior*, 2, 6-10. <https://doi.org/10.1038/s41562-017-0189-z>
- Bolton, A., & Newell, B.R. (2017). Applying behavioral science to government policy: Finding the 'Goldilocks zone'. *Journal of behavioral economics for policy*, 1, 9-14.
- Brandon, A., Clapp, C.M., List, J.A., Metcalfe, R.D., & Price, M. (2022). *The human perils of scaling smart technologies: evidence from field experiments*. (Working Paper No. 30482). National Bureau of Economic Research.
- Cairney, P. (2012). *Understanding public policy – Theories and issues*. Hampshire, United Kingdom: Palgrave Macmillan.

- Cairney, P. (2019). The UK government's imaginative use of evidence to make policy. *British Politics*, 14, 1-22. <https://doi.org/10.1057/s41293-017-0068-2>
- Cairney, P., & Oliver, K. (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health Research Policy Systems*, 15(35). <https://doi.org/10.1186/s12961-017-0192-x>
- Chametzky, B. (2016). Coding in classic grounded theory: I've done an interview; now what? *Sociology mind*, 6(4), 163-172.
- Congdon, W.J., Kling, J.R., & Mullainathan, S. (2011). *Policy and choice: Public finance through the lens of behavioral economics*. Washington D.C.: Brookings Institution Press.
- 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>
- Cowles, M., & Davis, C. (1982). On the origins of the .05 level of statistical significance. *American Psychologist*, 37(5), 553-558.
- Creswell, J.W. (2016). *30 essential skills for the qualitative researcher*. California: SAGE Publications Inc.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of economic literature*, 47(2), 315-375.
- DellaVigna, S., Woojin, K., & Linos, E. (2022). *Bottlenecks for Evidence Adoption*. (Working Paper No. 30144). National Bureau of Economic Research.
- Dewies, M., Denктаş, S., Giel, L., Noordzij, G., & Merkelbach, I. (2022). Applying behavioral insights to policy: An example from Rotterdam. *Global Implementation Research and Applications*, 2, 53-66.
- Feitsma, J. (2019a). Brokering behaviour change: the work of behavioural insights experts in government. *Policy & Politics*, 47(1), 37-56.
- Feitsma, J. (2019b). Meters maken: gedragsexperts in het lokaal bestuur. *Bestuurswetenschappen*, 73(1), 5-29.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. London: SAGE Publications Ltd.
- Haile, Z.T. (2023). Power analysis and exploratory research. *Journal of Human Lactation*, 39(4), 579-583. <https://doi.org/10.1177/08903344231195625>
- Harris, J.K. (2021). Primer on binary logistic regression. *Family Medicine and Community Health*, 9(1). DOI: 10.1136/fmch-2021-001290

- Head, B.W. (2015). Toward more ‘Evidence-Informed’ policy making? *Public Administration Review*, 76(3), 472-484. DOI: 10.1111/puar.12475
- Healy, L.M. (2006). *Logistic regression: an overview*. Eastern Michigan University, College of Technology.
- Heukelom, F., & Sent, E.M. (2017). Behavioral economics: From advising organizations to nudging individuals. *Journal of Behavioral Economics for Policy*, 1(1), 5-10.
- Hjort, J., Moreira, D., Rao, G., & Santini, J.F. (2019). *How research affects policy: experimental evidence from 2150 Brazilian municipalities*. (Working Paper No. 25941). National Bureau of Economic Research.
- Hoogerwerf, A. (2014). Beleid, processen en effecten. In A. Hoogerwerf & M. Herweijer (Red.), *Overheidsbeleid* (9e ed, pp. 15 – 28). Alphen aan den Rijn: Kluwer.
- Jamshed, S. (2014). Qualitative research method – interviewing and observation. *Journal of basic and clinical pharmacy*, 5(4), 87-88.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kamerstukken II, 34775 VI, nr. 88*. (2018, January 23th). Overheid.nl. Accessed on June 18<sup>th</sup> 2024, from <https://zoek.officielebekendmakingen.nl/kst-34775-VI-88.html>
- Keuzenkamp, H.A., & Magnus, J.R. (1995). On tests and significance in econometrics. *Journal of Economics*, 67, 5-24.
- Kim, J.H., & Choi, I. (2021). Choosing the level of significance: A decision-theoretic approach. *Abacus*, 57(1), 27-71.
- Kuehnhanss, C.R. (2019). The challenges of behavioral insights for effective policy design. *Policy and Society*, 38(1), 14-40. <https://doi.org/10.1080/14494035.2018.1511188>
- Lodge, M., & Wegrich, K. (2016). The rationality paradox of nudge: rational tools of a government in a world of bounded rationality. *Law & Policy*, 38(3), 250-267.
- Lourenço, J.S., Ciriolo, E., Almeida, S.R., & Troussard, X. (2016). *Behavioral insights applied to policy – European Report 2016*. Joint research centre. DOI: 10.2760/903938
- Lichand, G., & Serideira, A.P. (2023). *Behavioral insights for policy design: A new framework for understanding wicked social problems and designing policies for real citizens*. Switzerland: Springer.
- List, J.A. (2024). Optimally generate policy-based evidence before scaling. *Nature*, 626, 491-499. <https://doi.org/10.1038/s41586-023-06972-y>
- Lunn, P.D. (2012). Behavioral economics and policymaking: learning from the early adopters. *The economic and social review*, 43(3), 423-449.



- Lunn, P.D. (2014). *Regulatory Policy and Behavioral Economics*. Paris: OECD Publishing.
- Madrian, B.C. (2014). Applying insights from behavioral economics to policy design. *Annual review of economics* (6), 663-688. DOI: 10.1146/annurev-economics-080213-041033
- McChesney, F.S. Behavioral economics: Old wine in irrelevant new bottles? *Supreme Court Economic Review*, 21(1), 43-76.
- Michie, S., Van Stralen, M.M., & West, R. (2011). The behaviour change wheel: A new method for characterizing and designing behaviour change interventions. *Implementation Science* 6(42). <https://doi.org/10.1186/1748-5908-6-42>
- Milkman, K.L., Rogers, T., & Bazerman, M.H. (2010). I'll have the ice cream soon and the vegetables later: A study of online grocery purchases and order lead time. *Marketing Letters*, 21(1), 17-35.
- Miller, J., & Ulrich, R. (2019). The quest for an optimal alpha. *PLoS One*, 14(1). DOI: 10.1371/journal.pone.0208631
- Observatory of Public Sector Innovation. (May 10<sup>th</sup>, 2023). *Mapping the global behavioral insights community*. Accessed on March 7<sup>th</sup> from <https://oecd-opsi.org/blog/mapping-behavioural-insights/>
- Organisation for Economic Co-operation and Development. (2017). *Behavioral insights and public policy – Lessons from around the world*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2018). *Delivering better policies through behavioral insights: New approaches*. Paris: OECD Publishing.
- Palinkas, L.A., Hortwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., & Hoagwood, K. (2013). Purposeful sampling for qualitative data collection in mixed method implementation research. *Administration and policy in mental health and mental health services research*, 42(5), 533-544. DOI: 10.1007/s10488-013-0528-y
- Paradies, G., Klösters, M., Westbeek, E., Ooij, C., & Koning, N. (2023). *Het gebruik van gedragskennis voor beleidsontwikkeling – ervaren drijfveren, barrières en mogelijke oplossingsrichtingen bij het directoraat-generaal Klimaat en Energie*. (R11739). TNO.
- Ramos, T., & Matos, J. (2018). *RCTs are not (always) the answer*. Behavioral Scientist. Retrieved May 10<sup>th</sup>, 2024 from [RCTs Are Not \(Always\) the Answer - Behavioral Scientist](#)
- Raymaekers, P., & Brans, M. (2020). Nudging in perspectief: De verbreding van gedragsinzichten in beleid. *Beleidsonderzoek Online*, 1-28. DOI: 10.5553/BO/221335502020000004001

- Schiller, R.J. (2005). Behavioral economics and institutional innovation. *Southern Economic Journal*, 72(2), 268-283.
- Schreiber-Gregory, D.N., & Bader, K. (2018). *Logistic and linear regression assumptions: violation recognition and control*. (Conference paper).
- Schwartz, B. (2016). *The paradox of choice: why more is less*. New York: Ecco Press.
- Sent, E.M. (2004). Behavioral economics: how psychology made its (limited) way back into economics. *History of political economy*, 36(4), 735-760.
- Serdar, C.C., Cihan, M., Yücel, D., & Serdar, M.A. (2021). Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia medica*, 31(1). DOI: [10.11613/BM.2021.010502](https://doi.org/10.11613/BM.2021.010502)
- Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., and Aczél, B. (2017). A systematic scoping review of the choice architecture movement: Towards understanding when and why nudges work. *Journal of Behavioral Decision Making*.
- Thaler, R. H. (2016). Behavioral Economics: Past, Present and Future. *American Economic Review*, 106(7), 1577-1600. DOI: 10.1257/aer.106.7.1577
- Thaler, R. H. (2018). From cashews to nudges: The evolution of behavioral economics. *American Economic Review*, 108(6), 1265-1287. DOI: 10.1257/aer.108.6.1265
- Thaler, R.H., & Sunstein, C.R. (2008). *Nudge – Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.
- Troussard, X., & Van Bavel, R. (2018). How can behavioral insights be used to improve EU policy? *Intereconomics* 53, 8-12. <https://doi.org/10.1007/s10272-018-0711-1>
- Truc, A. (2021). Forty years of behavioral economics. *The European Journal of the History of Economic Thought*, 29(3), 393-437.
- Van Bavel, R. (2020). *Behavioral insights for EU policy making*. Brussels: Elsevier Limited.
- Van Bavel, R., Herrmann, B., Esposito, G., & Proestakis, A. (2013). *Applying behavioral sciences to EU policy-making*. Luxembourg: Publications Office of the European Union.
- Van de Riet, O., Pietersma, S., & Enserink, B. (2023). Gedragsinzichten bieden meer beleidskansen dan er nu worden benut. *ESB*, 108(4821), 236-237.
- Vivalt, E. & Coville, A. (2023). *How do policymakers update their beliefs?* World Bank.
- Wang, S. & Yang, D.Y. (2021). *Policy experimentation in China: The political economy of policy learning*. (Working Paper No. 29402).

- Weston, C., Gandell, T., Beauchamp, J., McAlpine, L., Wiseman., & Beauchamp, C. (2001). Analyzing interview data: The development and evolution of a coding system. *Qualitative Sociology*, 24(3), 381-400.
- Whitehead, M., Jones, R., Howell, R., Lilley, R., & Pykett, J. (2014). *Nudging all over the world - Assessing the global impact of the behavioural sciences on public policy*. United Kingdom: Economic and Social Research Council.
- Williams, M., & Moser, T. (2019). The art of coding and thematic exploration in qualitative research. *International Management Review*, 15(1), 45-55.
- Wilson VanVoorhis, C.R., & Morgan, B.L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in quantitative methods for psychology*, 3(2), 43-50.
- World Health Organization. (2023). *Use of behavioral and cultural insights in 2021-2022 in the WHO European Region: Status report*. Copenhagen: WHO Regional Office for Europe.

## Appendix I: Interview guide

Preface: First of all, thank you very much for participating in our research. We are Bente Hogema and Sam Wennink, master's students in Economics, Behavior, and Policy at Radboud University. The aim of this research is to gain insight into the application of behavioral insights within the government. The interview will take approximately one hour. To ensure that we capture all the information provided in this interview, we would like to record it. The recording will be used to transcribe the conversation, and it will be deleted afterward. You will remain anonymous in this research. Do you consent to this? If at any point you wish to withdraw, you are free to stop the interview. In that case, we will not use the interview in our research. Once the interview is transcribed, the transcript will be sent to you. You will then have the opportunity to read and review the text to ensure you stand by your words. You will also have the option to remove parts of the transcript or the entire transcript if desired. If you have any questions during the interview, or if you do not understand a question, please feel free to let us know. Do you have any questions before we begin?

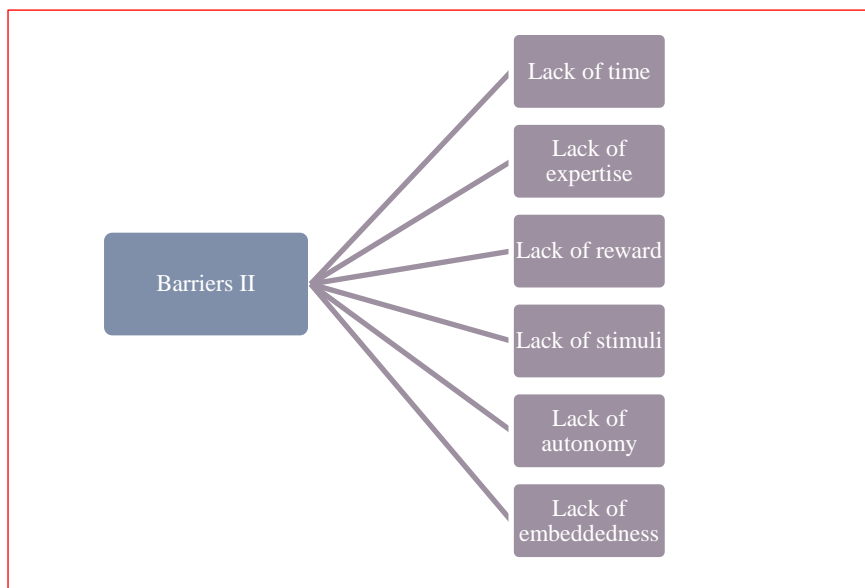
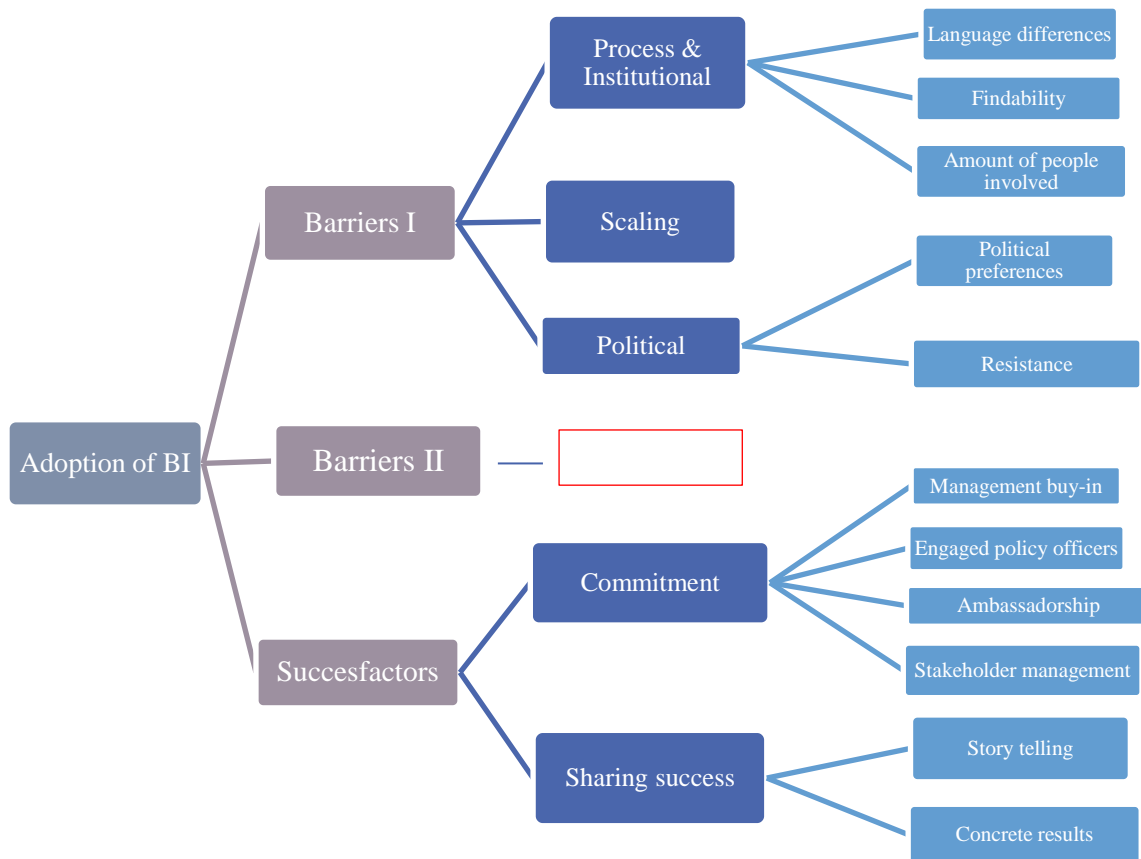
*Only the relevant questions for this study are listed below.*

1. Could you briefly tell us about your background? (Age, education, current position, and relevant tasks)
  - a. Do you work full-time or part-time?
  - b. Do you see yourself more as an expert/researcher or as an ambassador/initiator of using behavioural insights in your field of work?
2. You have contributed to a study submitted to BIN NL. To what extent do you feel that the impact of this study is representative of the impact achieved with the behavioral studies you generally contribute to?
  - a. Are there any other studies you have contributed to that were not published by BIN NL but still brought about a change in policy (i.e., implementation, supervision, or development of new policy)?
3. There are several points generally mentioned as barriers to using behavioural insights in the policy process. Examples include: lack of integration into the workflow, lack of encouragement from leadership, lack of time, lack of reward, lack of expertise, lack of space for choices. To what extent do you recognize these barriers?
  - a. Are there any barriers not mentioned that you do experience in your work?
4. The previously mentioned barriers relate to making evidence-based policy in general. Are there specific barriers you experience when implementing interventions (that have proven to be effective)?
5. Are there specific barriers you experience with regards of holding on to interventions that are implemented?
6. What do you consider to be the (most important) success factors for applying and maintaining behavioural insights derived from research in the development/implementation/enforcement of policy?
  - a. To what extent are these success factors present in practice?
7. How do you think the adoption and retention of behavioural insights in the development/implementation/enforcement of policy can be improved?
8. Since we aim to gather various perspectives on the topic, we would also like to interview policymakers/implementers/supervisors/internal clients. Do you know of any individuals we could approach who might be interested in participating?
9. Do you have any comments or questions?

## Appendix II: Overview of open codes

- (Te) Veel personen betrokken
- Aantonen belang
- Afstemming uitvoerder
- Ambassadeur op de goede plek
- Autonomie
- Averechts effect
- Bekendheid werken BI
- Betrokkenheid bij (hele) onderzoek
- Bewezen effectief genoeg
- Bewustwording
- Borging van kennis
- Brug OH-burger
- Budget
- Capaciteit beleidsdirectie
- Capaciteiten van beleidsmedewerkers
- Commitment manager
- Contact expert - medewerker
- Delen succesverhalen
- Doel om te landen in beleid
- Draagvlak
- Eigen taal spreken
- Evaluatie
- Expert vanaf begin betrekken
- Formaliseren
- Gebrek aan beloning
- Gebrek aan geld
- Gebrek aan inbedding
- Gebrek aan kennis
- Gebrek aan tijd
- Gedragscheck
- Gedragsinzicht gaat verloren in proces
- Geen autonomie voor keuze
- Geen politieke interesse
- Geen prioriteit
- Gefragmenteerde projecten
- Gevoelde urgentie
- Herkennen gedragsvraagstuk
- Hierarchie
- Hoge werkdruk
- Inertia
- Ingehaald realiteit
- Inrichting organisatie
- Inter-departementaal
- Kennis bij beslisser
- kennis bij uitvoerder
- Kennis verbreding
- Kennis verdwijnt
- Kennismakelaar
- Lastig onderzoek doen
- Lobby
- Local government
- Loophole in interventie
- Mate succesfactor aanwezig
- Mee met beleidsdirectie
- Meer dan vraaggestuurd werken
- Meer gedragsexperts
- Meetbaarheid resultaat
- Monitoring projecten
- Motivatie van medewerkers
- Observeren medewerkers
- Omgang ambtenaren
- Omvang gedragsteams
- Onbekendheid onderzoek
- Opdrachtgever
- Opschalen
- Past niet in traditonele instrumenten
- Politieke druk
- Politieke wind
- Prioriteit op de agenda
- Quote
- Representativiteit onderzoek
- Resultaten delen
- Risico mijding
- Samenwerking overheden
- Selectief onderzoek gebruik
- Stakeholders
- Stimulering leidinggevende
- Succesfactor
- Systeem als belemmering
- Tastbare resultaten
- Toepasbaarheid inzichten
- Vasthouden aan interventie
- Verankering
- Verdwijnen regeling
- Verlies impact
- Verloop op de werkvloer
- Vernieuwingen
- Vindbaarheid vd experts
- Vinden van de juiste persoon
- Vinden van info
- Weerstand
- Weerstand beleidsdirectie

**Appendix III: Code tree**



## **Appendix IV: Robustness checks**

There are certain assumptions of the logistic regression that must be met for a valid logistic regression analysis. These assumptions include appropriate outcome structure, observation independence and no multicollinearity among independent variables (Harris, 2021; Schreiber-Gregory & Bader, 2018). The assumption of linearity between predictors and the logit of the outcome variable does not apply, since the predictors in the model are not continuous (Field, 2018). These topics are discussed respectively in the following sections.

### **Outcome structure**

The assumption of an appropriate outcome structure requires the outcome variables to be fitting for the analysis. In other words, a binary logistic regression analysis assumes a dichotomous dependent variable. This assumption is met, as the dependent variable is uptake (yes/no).

### **Observation independence**

The assumption of observation independence requires – as the name gives away – that the observations should be independent from another. Put differently, ‘the observations should not originate from repeated measurements or matched data’ (Schreiber-Gregory & Bader, 2018). The data used in this study does to some extent violate the assumption. This is due to the fact that some of the respondents answered the survey for more than one project they submitted to BIN NL. In addition to this, there are a few studies that cover either the same topic or are follow-up studies on others, also (to some degree) violating the idea of independent observations.

### **Absence of multicollinearity**

An additional assumption is the absence of multicollinearity, referring to ‘a strong correlation between two or more predictors’ (Field, 2018). In the analysis, all the predictors are considered individually but ideally they would all be incorporated in the same model. To see if there are any correlations that could potentially violate this assumption, the collinearity statistics are composed. Tabel 5 shows the output. The output values of tolerance are all above the .1 and the VIF values are below 10, which are the threshold values for possible issues related to multicollinearity (Field, 2018).

Predictor	Tolerance	VIF
Aim of research	.751	1.331
Effectsize significance	.873	1.146
Use of controlgroup	.890	1.124
Type of organization	.766	1.306
Type of intervention	.817	1.224
Pre-existing communication	.874	1.144
Initiator of the project	.775	1.291

TABLE 5: COLLINEARITY DIAGNOSTICS

Table 5 is complemented with a cross table (Table X), which can also be used to interpret the mutual correlations. The cross table indicates that there are three significant correlations between predictors, that is between *initiator of the project* and *aim of research*, between *type of organization* and *aim of research*, between *type of organization* and *initiator of the project*. These correlations have values of approximately .3 or -.3 and hence indicate a medium effect (Field, 2018). If all predictors were considered in a collective model, these numbers combined with the output of Table 6 are not deemed strong enough to cause the need to remove predictors from the model.

		Aim of reserach	Effect size significance	Use of control-group	Type of organization	Type of intervention	Pre-existing communication	Initiator of the project
Aim of reserach	Correlation Coefficient	1,000	-0,123	0,146	-,357*	0,075	-0,045	,334*
	Sig. (2-tailed)		0,416	0,332	0,015	0,618	0,767	0,023
	N	46	46	46	46	46	46	46
Effect size significance	Correlation Coefficient	-0,123	1,000	-0,211	0,125	0,250	-0,048	-0,103
	Sig. (2-tailed)	0,416		0,160	0,406	0,094	0,752	0,496
	N	46	46	46	46	46	46	46
Use of controlgroup	Correlation Coefficient	0,146	-0,211	1,000	0,026	0,020	0,197	0,086
	Sig. (2-tailed)	0,332	0,160		0,866	0,895	0,190	0,569
	N	46	46	46	46	46	46	46
Type of organization	Correlation Coefficient	-,357*	0,125	0,026	1,000	0,178	-0,011	-,369*
	Sig. (2-tailed)	0,015	0,406	0,866		0,237	0,941	0,012
	N	46	46	46	46	46	46	46
Type of intervention	Correlation Coefficient	0,075	0,250	0,020	0,178	1,000	0,202	-0,129
	Sig. (2-tailed)	0,618	0,094	0,895	0,237		0,179	0,393
	N	46	46	46	46	46	46	46
Pre-existing communication	Correlation Coefficient	-0,045	-0,048	0,197	-0,011	0,202	1,000	0,126
	Sig. (2-tailed)	0,767	0,752	0,190	0,941	0,179		0,405
	N	46	46	46	46	46	46	46
Initiator of the project	Correlation Coefficient	,334*	-0,103	0,086	-,369*	-0,129	0,126	1,000
	Sig. (2-tailed)	0,023	0,496	0,569	0,012	0,393	0,405	
	N	46	46	46	46	46	46	46

\*. Correlation is significant at the 0.05 level (2-tailed).

TABLE 6: CORRELATIONS OF THE INDEPENDENT VARIABLES



## Appendix V: Categorization of psychological biases

Behavioral insights draw upon a broad range of psychological biases, which can be categorized into three different categories. DellaVigna (2009) captures these biases – also known as deviations from the standard economic model – in nonstandard preferences, nonstandard beliefs and nonstandard decision making. Congdon, Kling & Mullainathan (2011) distinguish slightly different categories and classify deviations from the standard economic model in imperfect optimization, bounded self-control and nonstandard preferences. Imperfect optimization refers to individuals not being able to maximize their own welfare and inconsistencies in their choices. This contradicts the standard economic model assumption of utility maximization, which poses that individuals both have full knowledge over their preferences and can rank these preferences consistently. An example of this is the paradox of choice: increasing the number of choices for consumers decreases the likelihood of people actually proceeding to purchase something, while this should not affect the decision making process (or only in a positive way) according to the standard theory (Schwartz, 2016). Bounded self-control comprises the difficulty individuals experience when it comes to pursuing their intentions over time. This gap between intentions and actions is not captured by the standard model, which assumes time consistency in preferences. This can be illustrated by an example provided by (Milkman, Rogers & Bazerman, 2010) who find that people who get to choose between delivery of foods shortly or in a few days' time, opt for unhealthy foods for the short-term delivery but prefer healthy foods when it gets delivered in a few days. The third category, nonstandard preferences, captures how individuals' choices depend on their environment and take into account other people. Agents are susceptible to how choices are framed, the status quo and (other) reference points. Moreover, they comply to social norms and are inequity averse. This deviates from the standard model, which assumes preferences based on end-states and self-interest. An example from this category is people's tendency to stick to a default setting that is provided when making a decision (Congdon et al., 2011; Madrian, 2014).