

Rosanna Nagtegaal
Lars Tummers
Mirko Noordegraaf
Utrecht University

Victor Bekkers
Erasmus University

Designing to Debias: Measuring and Reducing Public Managers' Anchoring Bias

Research Article

Abstract: *Public managers' decisions are affected by cognitive biases. For instance, employees' previous year's performance ratings influence new ratings irrespective of actual performance. Nevertheless, experimental knowledge of public managers' cognitive biases is limited, and debiasing techniques have rarely been studied. Using a survey experiment on 1,221 public managers and employees in the United Kingdom, this research (1) replicates two experiments on anchoring to establish empirical generalization across institutional contexts and (2) tests a consider-the-opposite debiasing technique. The results indicate that anchoring bias replicates in a different institutional context, although effect sizes differ. Furthermore, a low-cost, low-intensity consider-the-opposite technique mitigates anchoring bias in this survey experiment. An exploratory subgroup analysis indicates that the effect of the intervention depends on context. The next step is to test this strategy in real-world settings.*

Evidence for Practice

- In survey experiments, anchoring bias is robust across institutional contexts for managers as well as employees. It should be considered when designing decision environments for public management practices.
- Anchoring bias can be reduced by using a low-cost, low-intensity version of the consider-the-opposite strategy. This strategy involves asking for two reasons why the anchor is inappropriate. The effectiveness of this strategy depends on context and should be tested in real-world settings.
- Consider-the-opposite could be effective in internally oriented management practices, such as goal setting and performance feedback.

Appropriate practices by public managers are essential for public sector performance and, as a consequence, for well-functioning bureaucracies (Favero, Meier, and O'Toole 2016). However, decisions made by all human agents are subject to predictable cognitive biases (Tversky and Kahneman 1974). Cognitive biases occur when "human cognition reliably produces representations that are systematically distorted compared to some aspect of objective reality" (Haselton, Nettle, and Murray 2015, 968). For instance, people make different decisions when information is framed negatively than when it is framed positively (Bellé, Cantarelli, and Belardinelli 2018; Tversky and Kahneman 1981).

Empirical evidence shows that public managers use cognitive biases in public sector decision-making (Battaglio et al. 2018). Availability bias and anchoring bias are important, for instance, in macroeconomic forecasts that provide policy information for managing the U.S. economy (Krause 2006), framing matters for performance evaluations of organizations and individuals (Belardinelli et al. 2018), and

local election officials' overconfidence in their own judgment affects their technology preferences (Moynihan and Lavertu 2012). Despite this recognition, the body of knowledge on the effects of cognitive biases on public managers' decision-making is still limited.

This article reports on a research strategy of initially replicating two experiments by Bellé, Cantarelli, and Belardinelli (2017, 2018) on cognitive biases in the public sector. These experiments represent two types of core internal management practices in the public sector: goal setting (e.g., establishing the maximum number of days within which to respond to emails) and performance feedback (e.g., performance ratings) (Favero, Meier, and O'Toole 2016; Pedersen and Stritch 2018a).

The aim of this replication is threefold (Jilke et al. 2017). First, the replication extends the generalizability of earlier results. Experimental results of one population and context might not generalize to another (Lykken 1968). That is why empirical generalization is essential to test the robustness of

Rosanna Nagtegaal is a doctoral candidate in the Utrecht School of Governance, Utrecht University, the Netherlands. In her research, she studies behavioral public administration in order to analyze the (changing) behavior of public sector employees and professional workers in public services, especially in digitalizing environments. She most often uses experimental methodology.
Email: r.nagtegaal@uu.nl

Lars Tummers is professor of public management and behavior in the Utrecht School of Governance, Utrecht University, the Netherlands. His main research interests are public management, stereotypes, leadership, and behavioral change. He is developing—with others—an interdisciplinary field combining psychology and public administration, called behavioral public administration.
Email: l.g.tummers@uu.nl

Mirko Noordegraaf is professor of public management in the Utrecht School of Governance, Utrecht University, the Netherlands. He focuses on public management issues, with a particular emphasis on reform, public managers, managing professionals, public professionals, and (changing) professionalism. He is the author of the textbook *Public Management: Performance, Professionalism and Politics* (Palgrave Macmillan, 2015).
Email: m.noordegraaf@uu.nl

Victor Bekkers is professor of public administration and public policy in the Department of Public Administration and Sociology at Erasmus University. He is also acting dean of the Erasmus School of Social and Behavioral Sciences. His research interests focus on the role of information communication technology and innovation in policy and governance processes, policy implementation and public service management, and policy learning and innovation in relation to behavioral change.
Email: bekkers@essb.eur.nl

Public Administration Review, Vol. 80, Iss. 4, pp. 565–576. © 2020 The Authors. *Public Administration Review* published by Wiley Periodicals LLC on behalf of The American Society for Public Administration. DOI: 10.1111/puar.13211.

findings (Walker, James, and Brewer 2017). Second, replications reduce the chance of false positives (Ioannidis 2018). Third, the influence of context might be tested through replications. Therefore, experiments should be tested across politico-administrative contexts.

In our case, most evidence indicates that anchoring bias is very robust across contexts (Furnham and Boo 2011). Nevertheless, new research has shown that the effect of biases might be dependent on institutional context (Christensen 2018b; Dudley and Xie 2020; Holm 2017). In other words, cognitive processes apply to individuals, but they happen within institutions (Jones 2017). This institutional perspective on the effects of cognitive processes, including biases, remains understudied.

Therefore, we are interested in how the effects of cognitive biases can be generalized to other institutional contexts. In this case, this research compares the original results from Italy with results from the United Kingdom, as these two countries represent different politico-administrative regimes (Pollitt and Bouckaert 2011). Replications remain rare in research even though the importance of replications is evident (Brandt et al. 2014; Pedersen and Stritch 2018b; Walker, James, and Brewer 2017). Thus, this article contributes to building a body of knowledge, rather than relying on one experiment to substantiate claims.

Further, this research focuses on testing a low-intensity, low-cost debiasing technique. The impact of biases on decision-making has led scholars to suggest that biases should be taken into account when designing the architecture of jobs and tasks (Bellé, Cantarelli, and Belardinelli 2018; Vaughn and Linder 2018). However, in general, research demonstrating biases is more widespread than research on solving bias-related problems (Bhanot and Linos 2020). It seems to be “more newsworthy to show that something is broken than to show how to fix it” (Larrick 2004, 334). An explanation for this is that cognitive biases are robust and debiasing is notoriously difficult (Lilienfeld, Ammirati, and Landfield 2009). Nevertheless, given the effects of biases, debiasing has the potential to increase public sector performance, and therefore it should be on the agendas of practitioners as well as researchers. This article is focused on testing a debiasing technique that can be easily applied in practice.

This article concentrates on one specific cognitive bias: anchoring bias. Anchoring bias refers to the tendency to estimate unknown quantities by using an initial value (Tversky and Kahneman 1974). This bias is chosen for two reasons. First, anchoring bias has real-world implications for public management. Examples are quantitative evaluations of employees’ performance, where last year’s employee ratings affect this year’s ratings (Bellé, Cantarelli, and Belardinelli 2017); decisions on academic promotions, where performance criteria inform decisions, irrespective of performance (Chen and Kemp 2015); negotiations, where initial offers anchor negotiation outcomes (Guthrie and Orr 2006); and evaluations, where political and historical performance results label current performance as either a failure or success (Holm 2017).

Second, anchoring is notoriously robust, and previous attempts to debias anchoring in public management have not succeeded (Cantarelli, Bellé, and Belardinelli 2018; Furnham and Boo 2011).

However, there is a promising strategy that could be translated into public sector practices: consider-the-opposite. Although this strategy has been tested in other settings and has proven to be successful (Lord, Lepper, and Preston 1984; Mussweiler, Strack, and Pfeiffer 2000), no reported experiments exist that test consider-the-opposite as a low-cost, low-intensity intervention to debias decisions in public management. This leads to the following research question: *Does anchoring bias affect public management decisions across institutional contexts, and can anchoring bias in decision-making be mitigated through a low-cost, low-intensity consider-the-opposite strategy?*

The article is structured as follows: First, it elaborates on the theoretical background by discussing cognitive biases and specifically anchoring bias. The article also expands on debiasing techniques and argues that a consider-the-opposite strategy is an appropriate strategy for debiasing anchoring effects for public management decisions. Second, the method is explained: a survey experiment involving 1,221 public managers and employees, part of which is a replication. Third, the authors discuss the results and, fourth, their implications for public management practice and scholarship are considered.

Theoretical Framework

First, the article explains the most important concepts used in this study.

Cognitive Biases

To understand cognitive biases, it is necessary to start with dual process theory. Dual process theory is a broad cognitive theory about the workings of the human mind and shows that people make decisions through two interconnected cognitive “systems” (Chaiken and Trope 1999; Evans and Stanovich 2013; Kahneman 2011). System 1 allows people to make decisions rapidly, automatically, and intuitively. System 1 is the oldest of the two systems and not exclusive to humans. System 1 is particularly useful in dangerous situations as it allows us to act without consciously having to think. The other system, system 2, is slower and more reflective. Applying system 2, people can go beyond our first hunch and consider more complex factors that may be relevant to the problem at hand.

In system 1 decisions, shortcuts, or heuristics, are used. Heuristics simplify complex decisions (Tversky and Kahneman 1974). In many situations, this is beneficial, but it can sometimes lead to systematic biases. Researchers have been particularly successful at identifying biases, with more than 175 biases detected so far (Benson 2016). Notable examples are status quo bias, which refers to the tendency of people to stick to the current situation (Kahneman, Knetsch, and Thaler 1991), and confirmation bias, which refers to people interpreting or looking for information that is concurrent with their existing beliefs (Nickerson 1998). Cognitive biases influence decisions made by public managers and employees (Battaglio et al. 2018; Gimmelikhuijsen et al. 2016).

Anchoring

This article focuses on anchoring bias. As noted earlier, anchoring bias has many real-world effects, including influencing performance ratings, and it is notoriously robust (Bellé, Cantarelli, and Belardinelli 2017; Furnham and Boo 2011). The current dominant

view is that anchoring works by activating anchor-consistent information (Furnham and Boo 2011). As such, anchoring is an association-based bias (Larrick 2004). Association-based biases make some information more available in the mind than other information during decision-making, creating selective accessibility (Strack and Mussweiler 1997). In essence, anchoring bias thus induces a reference frame in a person, a certain set of thoughts, which makes it difficult to consider alternative possibilities (Koehler 1991).

In this article, two earlier anchoring experiments are replicated (Bellé, Cantarelli, and Belardinelli 2017, 2018). These experiments represent two core categories of internal management practices (Favero, Meier, and O'Toole 2016; Pedersen and Stritch 2018a). Internally focused managerial practices aim to change employees' behavior. They fall into four broad categories: they set goals, build trust, increase participation in decision-making, or provide performance feedback.

The first experiment illustrates a goal-setting practice. Specifically, the scenario used here concerns establishing a maximum number of days that employees have to respond to citizens' emails (Bellé, Cantarelli, and Belardinelli 2018). Public organizations rely on written digital communication to send information to citizens and stakeholders (Faulkner et al. 2018). Nevertheless, decisions about what constitutes a timely response could be influenced by anchors that might be irrelevant (Bellé, Cantarelli, and Belardinelli 2018).

The second experiment focuses on providing feedback on public employees' performance through performance ratings (Bellé, Cantarelli, and Belardinelli 2017). Performance ratings are a common form of performance appraisal in the public sector and are widely studied in public administration. However, performance ratings have been shown to be prone to errors and biases (Tummers 2017). Anchoring bias has proven to be very robust across different contexts (Furnham and Boo 2011), and there is no reason to believe otherwise in the case of internal management practices across institutional contexts. This leads to the following hypothesis:

Hypothesis 1: Participants in the high-anchor replication groups will report estimates that are significantly higher than estimates from participants in low-anchor replication groups.

Debiasing

As debiasing is fairly novel in public administration research, an overview of the debiasing literature is provided. There are two main overarching debiasing categories (Croskerry, Singhal, and Mamede 2013b; Keren 1990; Larrick 2004; Soll, Milkman, and Payne 2015). One category includes strategies that *modify the environment*, which either makes the bias irrelevant or mitigates its effect. People could, for instance, use nudging or hold public employees accountable for their decisions (Aleksovska, Schillemans, and Grimmelikhuijsen 2019; Nagtegaal et al. 2019). The second category involves *modifying the decision maker*, which can be done by providing education on the bias at hand and the consequences that a bias might have and/or providing tools to mitigate the effect of the bias (Lilienfeld, Ammirati, and Landfield 2009). Strategies that modify the decision maker are grounded in a two-model system of reasoning that is related to dual process theory. This

model assumes that people first make an intuitive judgment about a situation with system 1 and that this judgment can be corrected by reflective and more effortful thinking with system 2 (Milkman, Chugh, and Bazerman 2009; Morewedge et al. 2015).

This article focuses on *modifying the decision maker* and the cognitive processes used to make a decision. Interventions that are informative and use system 2 have been argued to preserve individual dignity by allowing individual agency, and therefore they are preferred by the people affected by interventions (Sunstein 2016). Not all strategies that modify the decision maker are equally promising. For example, Christensen (2018a) shows that asking for justification decreases politicians' ability to make informed decisions. Cantarelli, Bellé, and Belardinelli (2018) tested an educational debiasing intervention in public service. This intervention did not overcome anchoring bias.

Others claim that a combination of education and tools is needed to achieve effective debiasing (Adame 2016; Morewedge et al. 2015; Wilson and Brekke 1994). This can be very intensive in terms of resources and undesirable in the public sector, which is characterized by low resources (Lipsky 2010). However, others claim that offering only corrective tools could be effective (Larrick 2004). Here, one strategy seems especially promising to debias anchoring: the consider-the-opposite strategy. This strategy is discussed next.

Consider-the-Opposite Strategies. First, it is important to realize that no debiasing strategy works on all types of biases (Croskerry, Singhal, and Mamede 2013a). In practice, biases work through different mechanisms, and some biases might work through multiple mechanisms at once (Larrick 2004). In this article, we attempt to mitigate anchoring bias by using a consider-the-opposite strategy because this strategy is a good fit for association-based biases such as anchoring. By its nature, anchoring creates a cognitive reference frame, making it hard to consider alternative thoughts (Koehler 1991). Consider-the-opposite is a technique to break through this frame and open the door to alternative reasoning (Mussweiler, Strack, and Pfeiffer 2000).

The consider-the-opposite strategy has been found to be effective in dealing with biases, such as confirmation bias (Anderson 1982; Hirt and Markman 1995), framing (Cheng, Wu, and Lin 2014), and the anchoring effect (Adame 2016; Lord, Lepper, and Preston 1984; Mussweiler, Strack, and Pfeiffer 2000). The consider-the-opposite approach is administered mostly by simply asking people to list reasons why the anchor value is inappropriate (Adame 2016; Kennedy 1995; Lord, Lepper, and Preston 1984; Mussweiler, Strack, and Pfeiffer 2000). In the past, consider-the-opposite has been tested, for example, on attitudes toward the death penalty, judging individuals' personality traits (Lord, Lepper, and Preston 1984), probabilities of a correct diagnosis (Arkes et al. 1988), estimating the value of a car, and estimating the probability of election outcomes (Mussweiler, Strack, and Pfeiffer 2000). Our experiment tests the consider-the-opposite approach on two scenarios representing two core internal management practices in the public sector. Moreover, our experiment uses real public managers and employees as a sample, increasing the external validity.

A further change from previous studies is that this research tests an online low-cost, low-intensity, and thus scalable, version of the consider-the-opposite strategy. Previously, most consider-the-opposite experiments have involved the researcher being present (Lord, Lepper, and Preston 1984; Mussweiler, Strack, and Pfeiffer 2000). However, recent research has shown that consider-the-opposite can work as an online intervention, provided that it is part of a training process (Adame 2016). The following hypotheses are formulated:

Hypothesis 2: Participants in the low-anchor consider-the-opposite group will report estimates that are higher than participants in the low-anchor replication group.

Hypothesis 3: Participants in the high-anchor consider-the-opposite group will report estimates that are lower than participants in the high-anchor replication group.

Methodology

Experiments frequently require a trade-off between control and internal validity and external validity and realism (Druckman et al. 2011). In this experiment, we opted for a controlled design with high internal validity. Therefore, we argue that if we do not find an effect here, we probably will not find an effect in more realistic scenarios. Our two scenarios are about establishing the maximum number of days within which an employee has to respond to citizens' inquiries (Bellé, Cantarelli, and Belardinelli 2018) and about giving a performance rating to an employee (Bellé, Cantarelli, and Belardinelli 2017). This research conducts replication experiments with the goal of achieving empirical generalization (Walker, James, and Brewer 2017). Consequently, the research design, measures, and analysis in the original experiment are used.

The original experiments were administered in Italy. This experiment is conducted on public managers and employees in the United Kingdom. The United Kingdom and Italy are interesting cases because they represent different politico-administrative regimes with different histories, rules, and practices. For instance, these countries differ in terms of many politico-administrative variables, including state structure and administrative culture (Pollitt and Bouckaert 2011). Italy is increasingly decentralized, whereas the United Kingdom has a more centralized structure. Apart from that, Italy has been described as a mild adopter of New Public Management (NPM) practices, which makes it a country in which managerial and traditional models are mixed (Nitzl, Sicilia, and Steccolini 2019). In contrast, the United Kingdom is a heavy adopter of NPM.

Specific to the scenarios, in the United Kingdom, the provision of information to citizens has been documented in the Freedom of Information Act (FOIA) of 2000 (Worthy 2010). The FOIA dictates a maximum of 20 days for responding to citizens' requests for most governmental organizations (Information Commissioner's Office 2019). In Italy, a FIOA was passed in 2016, establishing a maximum of 30 days to respond (Repubblica.it 2016). The original studies were conducted in Italy just after adoption of the FIOA law in June–July 2016 (Bellé, Cantarelli, and Belardinelli 2017, 2018).

Concerning the performance feedback scenario, both the United Kingdom and Italy use performance assessments and feedback as a

regular part of human resource practices (OECD 2012a, 2012b). Both countries use performance criteria such as interpersonal skills, activities undertaken, and improvement of competencies. In both countries, performance assessments are of high importance for remuneration and career advancement. As such, performance feedback is relevant in Italy as well as the United Kingdom. Nevertheless, differences also exist. As stated earlier, Italy and the United Kingdom differ in their adoption of NPM practices. Research has shown that NPM can affect the ways in which performance information is used (Nitzl, Sicilia, and Steccolini 2019). This makes the United Kingdom an interesting case with which to assess empirical generalization.

The study was conducted using the crowdsourcing tool Prolific. Crowdsourcing refers to the use people participating in an online environment to complete a variety of tasks (Sheehan 2018). The benefits of crowdsourcing are large-scale recruitment of participants in a short time, low costs, and access to a broader population. The downsides are a lack of control over the context in which the respondent takes the survey, loss of naivety, and possibly ethical problems because no set standards for payment exist (Palan and Schitter 2018; Shank 2016). Prolific, however, has been designed with the academic community in mind and therefore addresses these downsides, for instance, by not allowing researchers to pay less than an established minimum wage.

We used the prescreening option on Prolific to select people from the United Kingdom who are public employees. To get paid, the participants had to complete the whole study. We used the “forced response” option in Qualtrics so that participants could not continue with the survey unless the questions were answered. This resulted in 1,221 respondents who were randomized and 1,202 respondents who finished the whole survey. The percentage of missing data is thus very small (1.5 percent). Cases with missing values for either the grouping variable or the dependent variable were excluded from the analysis of that dependent variable.

Replications need a highly powered sample to confirm that the effect of the original study is significant (Brandt et al. 2014). The sample size was chosen based on a pilot experiment of the whole study, involving 16 people. Based on this pilot, a power analysis was conducted, which led to an estimation of 282 respondents needed per group to corroborate a one-tailed hypothesis for our smallest effect (Cohen's $d = 0.21$). This was in line with the sample size of the original experiments.

The debiasing intervention is based on earlier consider-the-opposite experiments (Adame 2016; Lord, Lepper, and Preston 1984; Mussweiler, Strack, and Pfeiffer 2000). Respondents were asked to state two reasons why the anchor is inappropriate. The direction in which the anchor was inappropriate was specified. In other words, if the anchor was too low, people were asked to explain why the anchor was too low. If the anchor was too high, people were asked to state why it was too high. Therefore, the intervention was simple, low cost, and low intensity, and it could be applied even when the researcher was not present. All scenarios and interventions are shown in the appendix.

Two considerations were important in designing our intervention. First, asking for a large number of reasons has been shown to be

countereffective, as participants who have trouble generating new reasons might come to the conclusion that the anchor was right all along (Larrick 2004). Sanna, Schwarz, and Stocker (2002) showed that requiring two reasons was effective in decreasing hindsight bias, while listing 10 reasons was not. Based on this, six informal consultations were conducted in which the scenarios were presented to academics and public employees. They were asked to state three reasons why the anchor was inappropriate. The rapidness of responses was considered. If the third reason was preceded by a pause, the difficulty of coming up with the third reason was discussed. In general, respondents took more time to generate the third reason and responded by saying that the first two were easy but the third reason was more difficult. Second, the direction in which the anchor was inappropriate was specified, as this has been shown to be more effective than asking people to list reasons in general (Chapman and Johnson 1999). Earlier research showed that justification is not sufficient (Belardinelli et al. 2018) and that respondents need to allow opposing thoughts in order for the intervention to work.

The experiment has four groups. The first and second groups replicate the experiments reported in Bellé, Cantarelli, and Belardinelli (2017) and Bellé, Cantarelli, and Belardinelli (2018), respectively, to determine the extent to which the biases they revealed exist in a U.K. context. These groups are labeled the *low* and *high anchor* replication groups. Our replications are registered under <https://osf.io/mye2h/> on the Open Science Framework and use the materials of the original authors as well as Brandt et al.'s (2014) replication recipe. The third and fourth groups test the effect of consider-the-opposite approaches to debias anchoring effects. These groups are labeled the *high* and *low consider-the-opposite* groups. The setup is shown in figure 1.

The dependent variables were the maximum number of days in which employees must respond to an email and the performance

score on a 0–100 scale. To check that our randomization was working correctly, we included managerial status, gender, industry of employment, educational background, and age in our experiment (Bellé, Cantarelli, and Belardinelli 2017, 2018).

Analysis

First, the high and low replication groups are compared with a *t*-test (testing hypothesis 1). The replication is successful if a significant difference is obtained in the same direction as the original trial. Effect sizes are compared using Borenstein et al. (2009) with a *Z*-test. Second, again using *t*-tests, the replication groups are compared with the corresponding high and low consider-the-opposite debiasing groups. Debiasing is effective if the mean of the low consider-the-opposite group is significantly higher than the low-anchor group and is obtained (hypothesis 2) and/or if the high consider-the-opposite group report significantly lower means than the high-anchor group (hypothesis 3). Significance levels are established at 0.017 and corrected by the Bonferroni correction per experiment (0.05/3). We also conducted an exploratory subgroup analysis of managers and employees. Data and materials are available on <https://osf.io/mye2h/>.

All effect sizes are calculated using Lakens (2013). The magnitude of the effect sizes is reported in accordance with Sawilowsky (2009). Sawilowsky (2009) expanded the reporting of effect sizes by including reports of very small *d* (.01), very large (1.2), and extremely large effects (2.0), in addition to effect sizes as developed by Cohen. He did this as a reaction to Cohen's warning about an inflexible approach to effect sizes, leading to Cohen's original values to become standards. Sawilowsky (2009) also wanted to describe more effect sizes as observed in reality.

Results

Our randomization checks showed no differences based on sector, manager, gender, or age. However, the checks did reveal a small

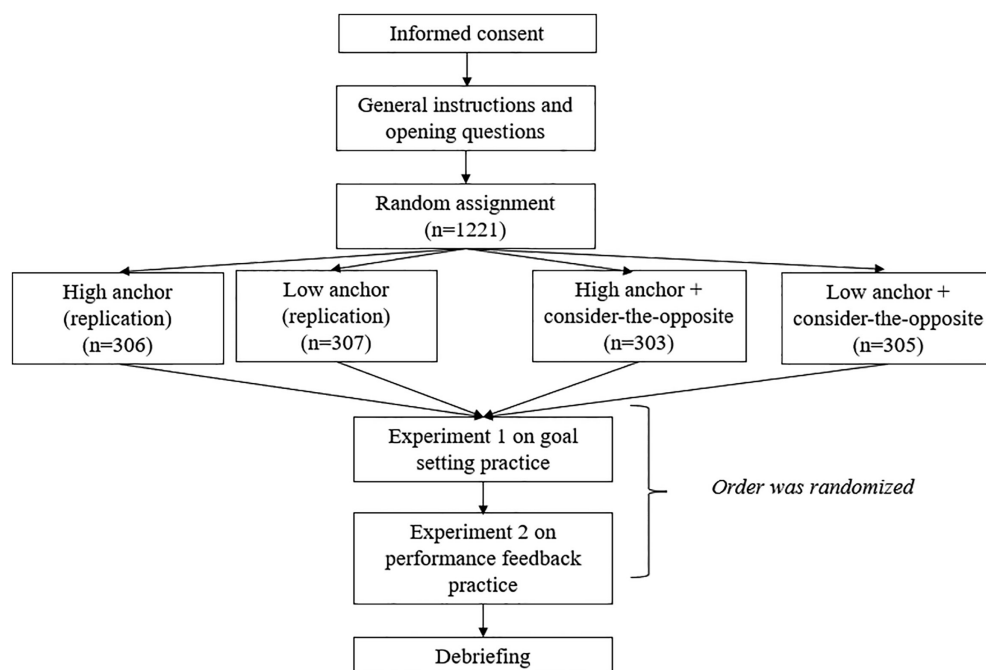


Figure 1 Experimental Design and Flow

difference between educational backgrounds. All the descriptives for each group are shown in table 1.

The results of experiments 1 and 2 support the first hypothesis, that participants in the high-anchor replication groups will report estimates that are significantly higher than the estimates from participants in the low-anchor replication groups. This indicates that anchors can affect managerial decisions concerning goal setting and performance feedback. In experiment 1, public managers and employees were asked to report the maximum number of days within which a public employee should reply to inquiries from citizens (replicating Bellé, Cantarelli, and Belardinelli 2018). Here, the replication was successful: the mean score for the low-anchor condition ($M = 3.69, SD = 2.98, N = 306$) was significantly lower ($t[316] = 16.93, p = .00$) than the score for the high-anchor condition ($M = 23.73, SD = 20.46, N = 305$). However, the effect size differs from the original study. The original study showed a medium effect size (Cohen's $d = 0.41$), whereas in the current study, the effect is very large (Cohen's $d = 1.37$). Furthermore, the mean minima and maxima differ from the original experiment. Bellé, Cantarelli, and Belardinelli (2018) reported a low-anchor mean of 31.82 and a high-anchor mean of 53.07. Our means (3.69, 23.73) indicate a shift of approximately 30 absolute points in the U.K. context. The means of the replication groups of experiment 1 are shown in figure 2.

Table 1 Descriptives and Differences per Group per Condition

	Low	CTO Low	CTO High	High	All
% female	80.5	78.4	78.2	78.1	78.8
% manager	32.2	30.2	32.7	29.7	31.2
Average age	39.41	39.34	39.19	38.88	39.21
<i>Sector of employment</i>					
% health care	23.5	22.6	25.1	22.5	23.4
% education	41.0	41.0	39.6	42.8	41.1
% administration	15.6	13.4	16.8	13.4	14.8
% other	19.9	23.0	18.5	21.2	20.6
<i>Educational background*</i>					
% technical and scientific degree	22.5	24.3	28.1	32.4	26.8
% social and humanities degree	45.9	48.2	46.9	48.4	47.3

Notes: The differences between groups were tested through chi-square tests apart from the difference in average age, which was calculated using an ANOVA. Significant differences (<0.05) are indicated by an asterisk. CTO = consider-the-opposite.

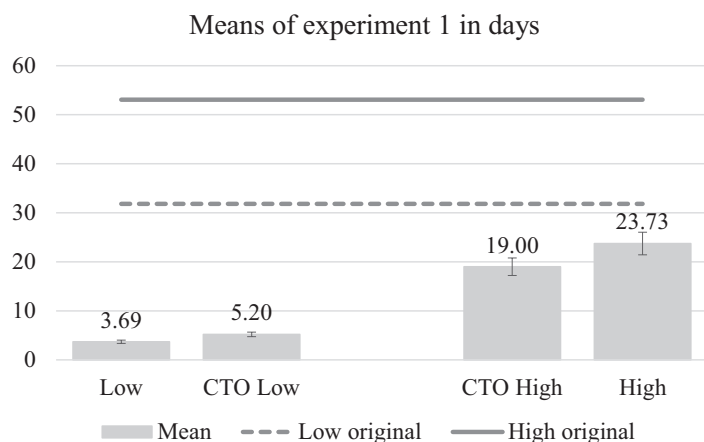


Figure 2 Means of Experiment 1 and Bellé, Cantarelli, and Belardinelli (2018) in Days

In experiment 2, public managers and employees were asked to rate the performance of an employee (replicating Bellé, Cantarelli, and Belardinelli 2017). In experiment 2, the replication was also successful: the mean rating for the low-anchor condition ($M = 69.69, SD = 10.18, N = 306$) was significantly lower ($t[595] = 24.27, p = .000$) than the rating for the high-anchor condition ($M = 88.24, SD = 8.68, N = 306$). In this experiment, the means were very similar to those in the original research in that Bellé, Cantarelli, and Belardinelli (2017) reported a low-anchor mean of 71.07 and a high-anchor mean of 88.47. Nevertheless, the effect size differed. In the original study, the effect size was very large (Cohen's $d = 1.21$). In this study, the effect could be labeled extremely large (Cohen's $d = 1.96$). The Z-tests provide Z-scores of -7.69 for the goal-setting experiment ($p = .000$) and -5.69 for the performance rating experiment ($p = .000$). This means that the effect sizes of the original studies differ significantly of those of the replication studies. The means of the replication groups for experiment 2 are shown in figure 2.

The second hypothesis, that participants in the low-anchor consider-the-opposite group will report estimates that are higher than participants in the low-anchor replication group, was also corroborated. This means that consider-the-opposite interventions can indeed debias goal-setting and performance feedback practices. Our consider-the-opposite intervention significantly increased ($M = 5.2, SD = 4.04, N = 298$) ($t[545] = 5.20, p = .000$) estimates from the low-anchor group in experiment 1 (low-anchor values $M = 3.69, SD = 2.98, N = 306$). The effect size is medium (Cohen's $d = 0.43$). Figure 2 shows the differences between the consider-the-opposite group and the low-anchor group in experiment 1. In experiment 2, our consider-the-opposite intervention significantly increased ($M = 71.73, SD = 7.17, N = 297$) ($t[549] = 2.85, p = .005$) estimates for the low-anchor group (low-anchor values $M = 69.69, SD = 10.18, N = 306$). This is a small effect size (Cohen's $d = 0.23$). Figure 3 shows the differences between the consider-the-opposite group and the low-anchor group in experiment 2.

Our results further support hypothesis 3, that participants in the high-anchor consider-the-opposite group will report lower estimates than participants in the high-anchor replication group. The consider-the-opposite intervention produced significantly lower ($t[567] = -3.192, p = .001$) estimates ($M = 19.00, SD = 15.59, N = 295$) than the high-anchor group (high-anchor values $M = 23.73, SD = 20.46, N = 305$) for experiment 1. This is a small effect size (Cohen's $d = 0.26$). The consider-the-opposite intervention also significantly lowered ($t[600] = -9.31, p = .000$) performance ratings ($M = 81.56, SD = 8.93, N = 296$) compared with the high-anchor group (high-anchor value $M = 88.24, SD = 8.68, N = 306$) for experiment 2. Here, the effect size is large (Cohen's $d = 0.76$). Figure 3 shows the results of experiment 2.

The sample consists of managers and employees. Even though we did not hypothesize differences between these groups beforehand, we conducted exploratory subgroup analyses. These analyses indicate that managers as well as employees were susceptible to anchors in both scenarios. For the consider-the-opposite intervention, we found that the effect depends on context. More specifically, we found that the consider-the-opposite intervention worked for the high anchor in the performance feedback scenario and the low anchor in the goal-setting scenario for managers. For

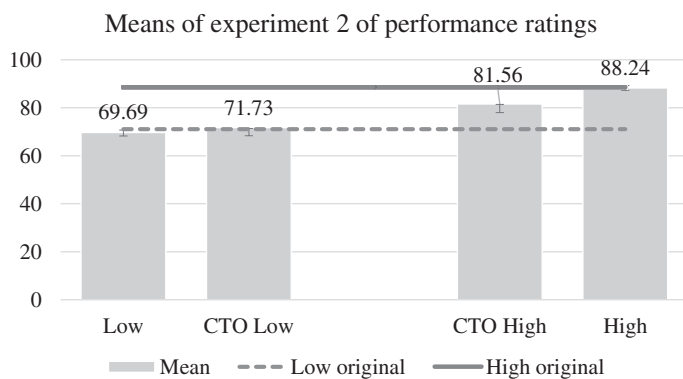


Figure 3 Means of Experiment 2 and Bellé, Cantarelli, and Belardinelli (2017) in Performance Ratings

the high-anchor goal-setting scenario, managers did not significantly adjust their reports while employees did. For the low-anchor performance rating, however, the opposite effect occurred, and managers did significantly adjust their anchors while the employees did not. In tables A1 and A2 in the appendix, all results of the exploratory subgroup analyses are shown.

Discussion

Recent research has shown that decisions by public managers are affected by cognitive biases. However, there is room to strengthen this body of knowledge, and, further, strategies to mitigate these biases are rarely studied. This article replicates two experiments representing two core internal management practices in a distinct institutional context. It shows how anchoring bias, one of the most robust of the biases identified, can be mitigated in the public sector by using a low-cost, low-intensity, and thus scalable, consider-the-opposite strategy. This article has several implications.

First, anchoring bias replicates across institutional contexts in our experiment. This empirically generalizes earlier findings (Bellé, Cantarelli, and Belardinelli 2017, 2018; Walker, James, and Brewer 2017). Nevertheless, statistical significance and effect direction only tell a part of the story. Effect sizes are vital when considering replications (Patil, Peng, and Leek 2016). Anchoring effects in the current study are significantly larger compared with the original study. For instance, for the experiment on goal setting, in the United Kingdom, a very large effect was found, compared with a medium effect in the original Italian study (Bellé, Cantarelli, and Belardinelli 2018). Furthermore, the mean minima and maxima shifted by about 30 absolute points compared with the original study. The differences in effects between Italy and the United Kingdom could be explained by multiple factors, including the survey, timing, and language. Our survey, for instance, exclusively focused on anchoring, while in the original experiments, the scenarios were part of a lengthier survey in which different biases were tested. Political reference levels also could have caused different effects for the goal-setting scenario (Holm 2017). In other words, existing anchors such as the maximum number of days required by FOIA laws might influence the effect of anchors. Political reference levels could give an indication of the “right” answer, limiting the influence of anchors as a cue of important information on decision-making (Furnham and Boo 2011). Apart from that, other characteristics of the law could have an effect. The U.K. FIOA law has been in place since 2000, while the Italy law was

accepted in 2016. While the law has been relatively successful in the United Kingdom, in Italy, the effects of the FOIA on citizens requests remain contested (Diritto di Sapere 2017; Worthy 2010). As such, the age of the laws, the knowledge of the laws and when to apply them, as well as compliance with laws might affect anchoring bias.

Second, our consider-the-opposite strategy achieved the desired outcome in all four cases. Nevertheless, the effect sizes ranged from small to large. This indicates that the effectiveness of consider-the-opposite strategies varies case by case. This research shows that consider-the-opposite is most effective in situations in which a high anchor is presented. In these cases, the anchor could be considered more extreme. Some authors claim that extreme anchors lead to a larger anchoring effect than anchors that are more reasonable (Furnham and Boo 2011). On top of that, our consider-the-opposite intervention does not fully remove the influence of anchoring, which may imply that anchoring bias is hard to remove completely with debiasing interventions such as the consider-the-opposite strategy. The latter could be explained by different reasons for a person to follow an anchor (Furnham and Boo 2011). For instance, people could perceive the anchor as being relevant to the problem at hand. Subsequently, consider-the-opposite might only target some of the causes.

Third, our subgroup analyses indicate that anchoring has an effect on both managers and employees. The effect of our consider-the-opposite intervention nevertheless depends on context. For instance, managers adjust goal setting when asked to listed reasons why the low anchor is inappropriate. Managers however do not adjust goal setting for the high anchor. On top of that, employees lower their judgment when asked to consider the opposite in the case of a high anchor and move toward a level comparable with the managers’ reports. This might indicate that managers already reflect more critically on the high anchor or have a clearer view of political reference levels, such as FIOA laws (Holm 2017). The opposite effect occurs for the performance feedback scenarios. In case of a low anchor, employees do not significantly adjust their ratings after consider-the-opposite, while managers do. For the high anchor, both groups adjust their ratings.

Fourth, as our research offers a successful low-cost, low-intensity intervention that can easily be applied in public services, the next step would be to translate this method to real public management practices. A couple of difficulties arise in doing so. First, determining the number of reasons to be provided is crucial, and this might differ in each case. This requires a case-by-case approach. Our general advice is to not require too many reasons as this can backfire (Sanna, Schwarz, and Stocker 2002). Nevertheless, provided an adequate number of reasons is established, a consider-the-opposite approach could possibly be institutionalized in formal and informal ways (Secunda 2012). This brings us to the second concern: asking managers and employees to formally write down reasons why the anchor is inappropriate might induce a sense of accountability, which could have an effect in itself (Aleksavska, Schillemans, and Grimmelikhuijsen 2019).

Fifth, the consider-the-opposite strategy could also be applied to counteract other association-based biases, such as confirmation bias and hindsight bias (Larrick 2004). Additionally, a consider-the-opposite approach could be adapted to other internal management practices in the public sector such as practices to build trust or

increase participation in decision-making (Favero, Meier, and O'Toole 2016; Pedersen and Stritch 2018a). Furthermore, possible applications are related to street-level decisions such as during client-employee interactions, where associations concerning client deservingness are known to play a role (Guthrie and Orr 2006; Jilke and Tummers 2018; Schafer and Schafer 2018).

Conclusions

This survey experiment investigated whether anchoring bias replicates in decisions representing internal management practices across institutional contexts, the relevance of the anchoring effect among managers and employees, and whether public managers and employees can be debiased by a low-cost, low-intensity consider-the-opposite intervention.

Limitations

Our research inevitably has limitations. The main limitation of this research is that the experiments focused on control and internal validity, as they were conducted through an online sample based on simplified fictional scenarios (Bouwman and Grimmelikhuisen 2016; Harrison et al. 2004). Real-world scenarios might differ and involve more information or more complexity. This could affect anchoring bias and the consider-the-opposite strategy on decision-making. The effect of anchoring bias and consider-the-opposite could, for instance, be smaller in a real-world scenario. Even though these limitations exist, our scenarios are externally relevant for two reasons. First, anchoring bias proves to be robust across a variety of experimental manipulations and contextual factors (Furnham and Boo 2011). Even if anchors are self-generated, people do not adjust their estimates sufficiently (Epley and Gilovich 2001). On top of that, increased understanding of the problem at hand does not discard anchoring bias. Experts are susceptible to anchoring bias, too (English, Mussweiler, and Strack 2006; Guthrie and Orr 2006). Our subgroup analyses also indicate that the anchoring effect is relevant for managers as well as employees. For the consider-the-opposite intervention, we have less empirical knowledge relating to the generalizability of the consider-the-opposite strategy. The subgroup analyses indicate that the effectiveness of the consider-the-opposite strategy depends on context. This subgroup analysis should be interpreted with caution, however, as it is correlational in nature and we lack power to detect small effects (Gerber and Green 2012).

Another limitation of our research could be the use of the term "bias," which implies the relationship of a cognitive shortcut to negative outcomes. This has generated criticisms by Gigerenzer and coauthors (Gigerenzer and Goldstein 1996; Gigerenzer et al. 2008; Gigerenzer 1991), for instance. The main criticism of Gigerenzer and colleagues is that cognitive biases are tools for humans that help decision-making instead of impairing it (see also Vis 2019). This relates to discussions on what constitutes rationality. In the tradition of Kahneman and Tversky, heuristics can lead to suboptimal decisions compared with a normative standard, oftentimes in line with expected utility theory. In the Gigerenzer tradition, heuristics allow people to make decisions that fit the environment. Although a thorough discussion of rationality relating to heuristics is beyond the scope of this article (discussions include Kahneman and Tversky 1996; Samuels, Stich, and Bishop 2002; Stanovich 2011), two important points of this discussion should be emphasized. First, heuristics are not inherently bad or good, but their merit depends on context (Gigerenzer and Brighton 2009;

Tversky and Kahneman 1974). Second, the discussion of heuristics is, in the end, a normative discussion that relies on philosophical questions (Hands 2014). Therefore, in this article, expected utility theory is assumed as a normative ideal. Scholars are invited to use other normative standards to interpret the results of this study.

We draw three main conclusions and propose directions for future research. First, anchoring is relevant across institutional contexts. Future research could explore whether there is an interaction with the effect of institutions, or more specifically rules and expectations, on the effect of cognitive biases. Second, anchoring bias has an effect on both managers and employees, as such being a manager does not remove anchoring bias. Third, a consider-the-opposite strategy can mitigate anchoring effects in our goal-setting and performance feedback scenarios. This strategy consists of requesting two reasons why the anchor is inappropriate. Effectiveness, however, depends on at least on two contextual factors. First, effects seem stronger when anchors are more extreme. Future research could also specify to what extent the perception of extremeness matter, for instance by asking whether respondents find specific anchors too low or too high. Second, managers might react differently to the consider-the-opposite strategy than employees. Managers for instance could have more knowledge of relevant laws. On the other hand, managers can be debiased in other cases. Future research could focus solely on managers for the aforementioned scenarios. This strategy has the potential to be used to address other association-based biases, as well as with other internal management practices and other types of employees such as street-level bureaucrats.

The practical implications of this research are that anchoring should be taken seriously in public management contexts, and could influence goal-setting and performance feedback practices. Therefore, it should be considered when designing jobs and tasks. Apart from that, our research indicates that in cases in which anchoring bias is a problem, a consider-the-opposite strategy is a promising tool to mitigate anchoring bias. This research should be seen as a first step toward mitigating anchoring bias. The next step is to test it in real-world scenarios. As this research has shown that anchoring bias can work differently depending on context and a consider-the-opposite technique mitigates anchoring bias in a controlled setting, we encourage scholars to test these results in more realistic settings. Future research should therefore focus on real-world scenarios prone to biases and field experiments to test the consider-the-opposite strategy.

Acknowledgments

We would like to thank our colleagues and peers at the Utrecht University School of Governance, NIG 2019, and EGPA 2019, as well as three anonymous reviewers for giving feedback on earlier versions of this article.

Funding

Lars Tummers acknowledges funding of NWO Grant 016.VIDI.185.017. Furthermore, he acknowledges that this work was supported by the National Research Foundation of Korea Grant, funded by the Korean Government (NRF-2017S1A3A2067636).

References

- Adame, Bradley J. 2016. Training in the Mitigation of Anchoring Bias: A Test of the Consider-the-Opposite Strategy. *Learning and Motivation* 53: 36–48. <https://doi.org/10.1016/J.LMOT.2015.11.002>.

- Aleksovska, Marija, Thomas Schillemans, and Stephan Grimmelikhuijsen. 2019. Lessons from Five Decades of Experimental and Behavioral Research on Accountability: A Systematic Literature Review. *Journal of Behavioral Public Administration* 2(2). <https://doi.org/10.30636/jbpa.22.66>.
- Anderson, Craig A. 1982. Inoculation and Counterexplanation: Debiasing Techniques in the Perseverance of Social Theories. *Social Cognition* 1(2): 126–39. <https://doi.org/10.1521/soco.1982.1.2.126>.
- Arkes, Hal R., David Faust, Thomas J. Guilmette, and Kathleen Hart. 1988. Eliminating the Hindsight Bias. *Journal of Applied Psychology* 73(2): 305–7. <https://doi.org/10.1037/0021-9010.73.2.305>.
- Battaglio, R. Paul, Jr., Paolo Belardinelli, Nicola Bellé, and Paola Cantarelli. 2018. Behavioral Public Administration ad fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations. *Public Administration Review* 79(3): 304–20. <https://doi.org/10.1111/puar.12994>.
- Belardinelli, Paolo, Nicola Bellé, Mariafrancesca Sicilia, and Ileana Steccolini. 2018. Framing Effects under Different Uses of Performance Information: An Experimental Study on Public Managers. *Public Administration Review* 78(6): 841–51. <https://doi.org/10.1111/puar.12969>.
- Bellé, Nicola, Paola Cantarelli, and Paolo Belardinelli. 2017. Cognitive Biases in Performance Appraisal: Experimental Evidence on Anchoring and Halo Effects with Public Sector Managers and Employees. *Review of Public Personnel Administration* 37(3): 275–94. <https://doi.org/10.1177/0734371X17704891>.
- . 2018. Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions. *Public Administration Review* 78(6): 828–40. <https://doi.org/10.1111/puar.12960>.
- Benson, Buster. 2016. Cognitive Bias Cheat Sheet. Better Humans, September 1. <https://betterhumans.coach.me/cognitive-bias-cheat-sheet-55a472476b18> [accessed May 2, 2020].
- Bhanot, Syon P., and Elizabeth Linos. 2020. Behavioral Public Administration: Past, Present, and Future. *Public Administration Review* 80(1): 168–71. <https://doi.org/10.1111/puar.13129>.
- Borenstein, Michael, Larry V. Hedges, Julian P.T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons.
- Bouwman, Robin, and Stephan Grimmelikhuijsen. 2016. Experimental Public Administration from 1992 to 2014. *International Journal of Public Sector Management* 29(2): 110–31. <https://doi.org/10.1108/IJPSM-07-2015-0129>.
- Brandt, Mark J., Hans IJzerman, Ap Dijksterhuis, Frank J. Farach, Jason Geller, Roger Giner-Sorolla, James A. Grange, Marco Perugini, Jeffrey R. Spies, and Anna Van't Veer. 2014. The Replication Recipe: What Makes for a Convincing Replication? *Journal of Experimental Social Psychology* 50: 217–24. <https://doi.org/10.1016/J.JESP.2013.10.005>.
- Cantarelli, Paola, Nicola Bellé, and Paolo Belardinelli. 2018. Behavioral Public HR: Experimental Evidence on Cognitive Biases and Debiasing Interventions. *Review of Public Personnel Administration* 40(1): 56–81. <https://doi.org/10.1177/0734371X18778090>.
- Chaiken, Shelly, and Yaacov Trope. 1999. *Dual-Process Theories in Social Psychology*. New York: Guilford Press.
- Chapman, Gretchen B., and Eric J. Johnson. 1999. Anchoring, Activation, and the Construction of Values. *Organizational Behavior and Human Decision Processes* 79(2): 115–53. <https://doi.org/10.1006/obhd.1999.2841>.
- Chen, Zhe, and Simon Kemp. 2015. Anchoring Effects in Simulated Academic Promotion Decisions: How the Promotion Criterion Affects Ratings and the Decision to Support an Application. *Journal of Behavioral Decision Making* 28(2): 137–48. <https://doi.org/10.1002/bdm.1838>.
- Cheng, Fei-Fei, Chin-Shan Wu, and Hsin-Hui Lin. 2014. Reducing the Influence of Framing on Internet Consumers' Decisions: The Role of Elaboration. *Computers in Human Behavior* 37: 56–63. <https://doi.org/10.1016/J.CHB.2014.04.015>.
- Christensen, Julian. 2018a. Do Justification Requirements Reduce Motivated Reasoning in Politicians' Evaluation of Policy Information? An Experimental Investigation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3295014 [accessed May 2, 2020].
- . 2018b. *Let's Look at the Facts: An Investigation of Psychological Biases in Policymakers' Interpretation of Policy-Relevant Information*. PhD diss. Aarhus University.
- Croskerry, Pat, Geeta Singhal, and Silvia Mamede. 2013a. Cognitive Debiasing 1: Origins of Bias and Theory of Debiasing. *BMJ Quality & Safety* 22: ii58–64. <https://doi.org/10.1136/bmjqs-2012-001712>.
- . 2013b. Cognitive Debiasing 2: Impediments to and Strategies for Change. *BMJ Quality & Safety* 22: ii65–72. <https://doi.org/10.1136/bmjqs-2012-001713>.
- Diritto di Sapere. 2017. Ignoranza Di Stato—Rapporto Sull'applicazione Del Foia Italiano. <https://blog.dirittoadisapere.it/rapporto-foia/> [accessed November 14, 2019].
- Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia, eds. 2011. *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.
- Dudley, Susan E., and Zhouan Xie. 2020. Designing a Choice Architecture for Regulators. *Public Administration Review* 80(1): 151–6. <https://doi.org/10.1111/puar.13112>.
- Englich, Birte, Thomas Mussweiler, and Fritz Strack. 2006. Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making. *Personality and Social Psychology Bulletin* 32(2): 188–200. <https://doi.org/10.1177/0146167205282152>.
- Epley, N., and Gilovich, T. 2001. Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science* 12(5): 391–396.
- Evans, Jonathan St.B.T., and Keith E. Stanovich. 2013. Dual-Process Theories of Higher Cognition. *Perspectives on Psychological Science* 8(3): 223–41. <https://doi.org/10.1177/1745691612460685>.
- Faulkner, Nicholas, Kim Borg, Peter Bragge, Jim Curtis, Eraj Ghafoori, Denise Goodwin, Bradley S. Jorgensen, et al. 2018. The INSPIRE Framework: How Public Administrators Can Increase Compliance with Written Requests Using Behavioral Techniques. *Public Administration Review* 79(1): 125–35. <https://doi.org/10.1111/puar.13004>.
- Favero, Nathan, Kenneth J. Meier, and Laurence J. O'Toole, Jr. 2016. Goals, Trust, Participation, and Feedback: Linking Internal Management with Performance Outcomes. *Journal of Public Administration Research and Theory* 26(2): 327–43. <https://doi.org/10.1093/jopart/muu044>.
- Furnham, Adrian, and Hua Chu Boo. 2011. A Literature Review of the Anchoring Effect. *Journal of Socio-Economics* 40(1): 35–42. <https://doi.org/10.1016/J.SOCEC.2010.10.008>.
- Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.
- Gigerenzer, Gerd. 1991. How to Make Cognitive Illusions Disappear: Beyond “Heuristics and Biases.” *European Review of Social Psychology* 2(1): 83–115. <https://doi.org/10.1080/14792779143000033>.
- Gigerenzer, Gerd, and Daniel G. Goldstein. 1996. Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review* 103(4): 650.
- Gigerenzer, Gerd, Ralph Hertwig, Ulrich Hoffrage, and Peter Sedlmeier. 2008. Cognitive illusions reconsidered. *Handbook of experimental economics results*, 1, 1018–1034. [https://doi.org/10.1016/S1574-0722\(07\)00109-6](https://doi.org/10.1016/S1574-0722(07)00109-6).
- Gigerenzer, Gerd, and Henry Brighton. 2009. Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science* 1(1): 107–43. <https://doi.org/10.1111/j.1756-8765.2008.01006.x>.
- Grimmelikhuijsen, Stephan, Sebastian Jilke, Asmus Leth Olsen, and Lars Tummers. 2016. Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review* 77(1): 45–56. <https://doi.org/10.1111/puar.12609>.
- Guthrie, Chris, and Dan Orr. 2006. Anchoring, Information, Expertise, and Negotiation: New Insights from Meta-Analysis. *Ohio State Journal on Dispute Resolution* 21(3): 597–628.

- Hands, D. Wade. 2014. Normative Ecological Rationality: Normative Rationality in the Fast-and-Frugal-Heuristics Research Program. *Journal of Economic Methodology* 21(4): 396–410. <https://doi.org/10.1080/1350178X.2014.965907>.
- Harrison, Glenn W., John A. List, Stephen Burks, Colin Camerer, Jeffrey Carpenter, Shelby Gerking, R Mark Isaac, et al., 2004. Field Experiments. *Journal of Economic Literature* 42(4): 1009–55. <https://pubs.aeaweb.org/doi/pdfplus/10.1257/0022051043004577>.
- Haselton, Martie G., Daniel Nettle, and Damian R. Murray. 2015. The Evolution of Cognitive Bias. In *The Handbook of Evolutionary Psychology*, edited by David M. Buss, 968–87. Hoboken: Wiley. <https://doi.org/10.1002/9781119125563.evpsych241>.
- Hirt, Edward R., and Keith D. Markman. 1995. Multiple Explanation: A Consider-an-Alternative Strategy for Debiasing Judgments. *Journal of Personality and Social Psychology* 69(6): 1069–86. <https://doi.org/10.1037/0022-3514.69.6.1069>.
- Holm, Jakob Majlund. 2017. Double Standards? How Historical and Political Aspiration Levels Guide Managerial Performance Information Use. *Public Administration* 95(4): 1026–40. <https://doi.org/10.1111/padm.12379>.
- Information Commissioner's Office. 2019. What Should We Do When We Receive a Request for Information? <https://ico.org.uk/for-organisations/guide-to-freedom-of-information/receiving-a-request/> [accessed May 2, 2020].
- Ioannidis, John P.A. 2018. Why Most Published Research Findings Are False. In *Getting to Good: Research Integrity in the Biomedical Sciences*, edited by Arthur L. Caplan and Barbara K. Redman, 2–8. Cham, Switzerland: Springer International.
- Jilke, Sebastian, Nicolai Petrovsky, Bart Meuleman, and Oliver James. 2017. Measurement Equivalence in Replications of Experiments: When and Why It Matters and Guidance on How to Determine Equivalence. *Public Management Review* 19(9): 1293–310. <https://doi.org/10.1080/14719037.2016.1210906>.
- Jilke, Sebastian, and Lars Tummars. 2018. Which Clients Are Deserving of Help? A Theoretical Model and Experimental Test. *Journal of Public Administration Research and Theory* 28(2): 226–38. <https://doi.org/10.1093/jopart/muy002>.
- Jones, Bryan D. 2017. Behavioral Rationality as a Foundation for Public Policy Studies. *Cognitive Systems Research* 43: 63–75. <https://doi.org/10.1016/j.COGLYS.2017.01.003>.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. London: Penguin.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1991. Anomalies the Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives* 5(1): 193–206.
- Kahneman, Daniel, and Amos Tversky. 1996. On the Reality of Cognitive Illusions. *Psychological Review* 103(3): 582–591.
- Kennedy, Jane. 1995. Debiasing the Curse of Knowledge in Audit Judgment. *Accounting Review* 70(2): 249–73.
- Keren, Gideon. 1990. Cognitive Aids and Debiasing Methods: CAN Cognitive Pills Cure Cognitive Ills? *Advances in Psychology* 68: 523–52. [https://doi.org/10.1016/S0166-4115\(08\)61341-2](https://doi.org/10.1016/S0166-4115(08)61341-2).
- Koehler, Derek J. 1991. Explanation, Imagination, and Confidence in Judgment. *Psychological Bulletin* 110(3): 499–519.
- Krause, George A. 2006. Beyond the Norm. *Rationality and Society* 18(2): 157–91. <https://doi.org/10.1177/1043463106063322>.
- Lakens, Daniël. 2013. Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-Tests and ANOVAs. *Frontiers in Psychology* 4: 863. <https://doi.org/10.3389/fpsyg.2013.00863>.
- Larrick, Richard P. 2004. Debiasing. In *Blackwell Handbook of Judgment and Decision Making*, edited by Derek J. Koehler and Nigel Harvey, 316–38. Malden, MA: Blackwell. <https://doi.org/10.1002/9780470752937.ch16>.
- Lilienfeld, Scott O., Rachel Ammirati, and Kristin Landfield. 2009. Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspectives on Psychological Science* 4(4): 390–8. <https://doi.org/10.1111/j.1745-6924.2009.01144.x>.
- Lipsky, Michael. 2010. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. 30th anniversary. New York: Russell Sage Foundation.
- Lord, Charles G., Mark R. Lepper, and Elizabeth Preston. 1984. Considering the Opposite: A Corrective Strategy for Social Judgment. *Journal of Personality and Social Psychology* 47(6): 1231–43. <https://doi.org/10.1037/0022-3514.47.6.1231>.
- Lykken, David T. 1968. Statistical Significance in Psychological Research. *Psychological Bulletin* 70(3, Pt. 1): 151–9. <https://doi.org/10.1037/h0026141>.
- Milkman, Katherine L., Dolly Chugh, and Max H. Bazerman. 2009. How Can Decision Making Be Improved? *Perspectives on Psychological Science* 4(4): 379–83. <https://doi.org/10.1111/j.1745-6924.2009.01142.x>.
- Morewedge, Carey K., Haewon Yoon, Irene Scopelliti, Carl W. Symborski, James H. Korris, and Karim S. Kassam. 2015. Debiasing Decisions. *Policy Insights from the Behavioral and Brain Sciences* 2(1): 129–40. <https://doi.org/10.1177/12372732215600886>.
- Moynihan, Donald P., and Stéphane Lavertu. 2012. Cognitive Biases in Governing: Technology Preferences in Election Administration. *Public Administration Review* 72(1): 68–77. <https://doi.org/10.1111/j.1540-6210.2011.02478.x>.
- Mussweiler, Thomas, Fritz Strack, and Tim Pfeiffer. 2000. Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility. *Personality and Social Psychology Bulletin* 26(9): 1142–50. <https://doi.org/10.1177/01461672002611010>.
- Nagtegaal, Rosanna, Lars Tummars, Mirko Noordegraaf, and Victor Bekkers. 2019. Nudging Healthcare Professionals towards Evidence-Based Medicine: A Systematic Scoping Review. *Journal of Behavioral Public Administration* 2(2): 1–20. <https://doi.org/10.30636/jbpa.22.71>.
- Nickerson, Raymond S. 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* 2(2): 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>.
- Nitzl, Christian, MariaFrancesca Sicilia, and Ileana Steccolini. 2019. Exploring the Links between Different Performance Information Uses, NPM Cultural Orientation, and Organizational Performance in the Public Sector. *Public Management Review* 21(5): 686–710. <https://doi.org/10.1080/14719037.2018.1508609>.
- Organisation for Economic Co-operation and Development (OECD). 2012a. Human Resources Management Country Profiles: Italy. <https://www.oecd.org/gov/pem/OECD%20HRM%20Profile%20-%20Italy.pdf> [accessed May 5, 2020].
- Organisation for Economic Co-operation and Development (OECD). 2012b. Human Resources Management Country Profiles: United Kingdom. <https://www.oecd.org/gov/pem/OECD%20HRM%20Profile%20-%20United%20Kingdom.pdf> [accessed May 5, 2020].
- Palan, Stefan, and Christian Schitter. 2018. Prolific.ac—A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance* 17: 22–7. <https://doi.org/10.1016/j.JBEF.2017.12.004>.
- Patil, Prasad, Roger D. Peng, and Jeffrey T. Leek. 2016. What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. *Perspectives on Psychological Science* 11(4): 539–44. <https://doi.org/10.1177/1745691616646366>.
- Pedersen, Mogens Jin, and Justin M. Stritch. 2018a. Internal Management and Perceived Managerial Trustworthiness: Evidence from a Survey Experiment. *American Review of Public Administration* 48(1): 67–81. <https://doi.org/10.1177/0275074016657179>.
- . 2018b. RNICE Model: Evaluating the Contribution of Replication Studies in Public Administration and Management Research. *Public Administration Review* 78(4): 606–12. <https://doi.org/10.1111/puar.12910>.
- Pollitt, Christopher, and Geert Bouckaert. 2011. *Public Management Reform: A Comparative Analysis*. Oxford: Oxford University Press.
- Repubblica.it. 2016. Ecco Il Freedom of Information Act. *La Versione Definitiva*. May 19. http://www.repubblica.it/tecnologia/sicurezza/2016/05/19/news/ecco_

- il_freedom_of_information_act_la_versione_definitiva-140129607/ [accessed May 5, 2020].
- Samuels, Richard, London Stephen Stich, and Michael Bishop. 2002. *Ending the Rationality Wars: How to Make Disputes about Human Rationality Disappear*. Oxford: Oxford University Press.
- Sanna, Lawrence J., Norbert Schwarz, and Shevaun L. Stocker. 2002. When Debiasing Backfires: Accessible Content and Accessibility Experiences in Debiasing Hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(3): 497–502. <https://doi.org/10.1037/0278-7393.28.3.497>.
- Sawilowsky, Shlomo S. 2009. New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods* 8(2): 597–9.
- Schafer, Brad A., and Jennifer Kahle Schafer. 2018. Client Likeability in Auditor Fraud Risk Judgments: The Mitigating Influence of Task Experience, the Review Process, and a “Consider the Opposite” Strategy. *Current Issues in Auditing* 12(1): P11–6. <https://doi.org/10.2308/ciia-52118>.
- Secunda, Paul M. 2012. Cognitive Illiberalism and Institutional Debiasing Strategies. *San Diego Law Review* 49(2): 373–414.
- Shank, Daniel B. 2016. Using Crowdsourcing Websites for Sociological Research: The Case of Amazon Mechanical Turk. *American Sociologist* 47(1): 47–55. <https://doi.org/10.1007/s12108-015-9266-9>.
- Sheehan, Kim Bartel. 2018. Crowdsourcing Research: Data Collection with Amazon’s Mechanical Turk. *Communication Monographs* 85(1): 140–56. <https://doi.org/10.1080/03637751.2017.1342043>.
- Soll, Jack B., Katherine L. Milkman, and John W. Payne. 2015. A User’s Guide to Debiasing. In *Blackwell Handbook of Judgment and Decision Making*, edited by Derek J. Koehler and Nigel Harvey, 924–51. Malden, MA: Wiley. <https://doi.org/10.1002/9781118468333.ch33>.
- Stanovich, K. 2011. *Rationality and the Reflective Mind*. Oxford: Oxford University Press.
- Strack, Fritz, and Thomas Mussweiler. 1997. Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility. *Journal of Personality and Social Psychology* 73(3): 437–46.
- Sunstein, Cass R. 2016. People Prefer System 2 Nudges (Kind of). *Duke Law Journal* 66(1): 121–68.
- Tummers, Lars. 2017. The Relationship between Coping and Job Performance. *Journal of Public Administration Research and Theory* 27(1): 150–62. <https://doi.org/10.1093/jopart/muw058>.
- Tversky, Amos, and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185(4157): 1124–31.
- . 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211(4481): 453–8.
- Vaughn, Valerie M., and Jeffrey A. Linder. 2018. Thoughtless Design of the Electronic Health Record Drives Overuse, but Purposeful Design Can Nudge Improved Patient Care. *BMJ Quality & Safety* 27: 583–6. <https://doi.org/10.1136/bmjqs-2017-007578>.
- Vis, Barbara. 2019. Heuristics and Political Elites’ Judgment and Decision-Making. *Political Studies Review* 17(1): 41–52. <https://doi.org/10.1177/1478929917750311>.
- Walker, Richard M., Oliver James, and Gene A. Brewer. 2017. Replication, Experiments and Knowledge in Public Management Research. *Public Management Review* 19(9): 1221–34. <https://doi.org/10.1080/14719037.2017.1282003>.
- Wilson, Timothy D., and Nancy Brekke. 1994. Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations. *Psychological Bulletin* 116(1): 117–42. <https://doi.org/10.1037/0033-2909.116.1.117>.
- Worthy, Ben. 2010. More Open but Not More Trusted? The Effect of the Freedom of Information Act 2000 on the United Kingdom Central Government. *Governance* 23(4): 561–82. <https://doi.org/10.1111/j.1468-0491.2010.01498.x>.

Appendix—Conditions for Each Experiment

Experiment 1: Bellé, Cantarelli, and Belardinelli (2018).

You are the senior manager of the Public Relations Office in a medium-sized municipality. You have to decide the maximum number of days by which your subordinates have to reply to citizens' inquiries sent via emails. Consider whether the maximum number of days to reply to citizens' emails must be higher or lower than 2[90] working days.

Consider-the-opposite.

To make this decision, please first list two reasons why 2[90] days might be too short[long] to respond to citizens' emails.

- Reason 1 _____
- Reason 2 _____

Now, indicate the maximum number of days to reply to citizens' emails below.

Experiment 2: Bellé, Cantarelli, and Belardinelli (2017).

Imagine that you have to assess this year's performance of a subordinate of yours. During this year, your subordinate met the majority of goals, had good interpersonal skills with her colleagues, and showed moderate creativity in proposing new ideas for the improvement of the services.

The previous year, you assigned your subordinate a performance rating of 51/100[91/100]. You have to decide whether to assign a rating lower or higher than 51/100[91/100].

Consider-the-opposite.

To make this decision, please first list two reasons why 51/100[91/100] might be too low[high] a rating for the *employee's performance this year*.

- Reason 1 _____
- Reason 2 _____

Now, indicate how would you assess your subordinate on a scale from 0–100:

Results of Exploratory Subgroup Analysis

Table A1 Subgroup Analysis Goal-Setting Experiment

	Managers			Employees		
	Replication	Low CTO	High CTO	Replication	Low CTO	High CTO
t-score	9.29***	3.39**	-.37	14.55***	4.17***	-3.39**
Cohen's <i>d</i>	1.42	0.50	.00	1.39	0.41	0.33
<i>n</i>	189	186	184	422	418	416
M(SD)						
Low	3.42 (2.13)	3.42 (2.13)		3.82 (3.30)	3.82 (3.30)	
High	18.80 (15.58)		18.80 (15.58)	25.80 (21.89)		25.80 (21.89)
Low CTO		4.69 (2.94)			5.41 (4.41)	
High CTO			17.97 (15.25)			19.48 (15.77)

*** $p < .001$; ** $p < .01$.

Table A2 Subgroup Analysis Performance Feedback Experiment

	Managers			Employees		
	Replication	Low CTO	High CTO	Replication	Low CTO	High CTO
t-score	16.94***	2.58*	-6.95***	18.26***	1.63	-6.58***
Cohen's <i>d</i>	2.47	0.38	1.02	1.72	0.16	0.65
<i>n</i>	190	187	186	422	416	416
M(SD)						
Low	67.58 (9.46)	67.58 (9.46)		70.71 (10.38)	70.71 (10.38)	
High	88.36 (7.18)		88.36 (7.18)	88.20 (9.25)		88.20 (9.25)
Low CTO		70.77 (7.14)			72.13 (7.17)	
High CTO			79.61 (9.74)			82.49 (8.39)

*** $p < .001$; ** $p < .01$; * $p < .05$.