

**Jasper Eshuis**  
Erasmus University Rotterdam  
**Thijs van de Geest**  
Rabobank  
**Erik Hans Klijn**  
Erasmus University Rotterdam

**Joris Voets**  
Ghent University  
**Magdalena Florek**  
Poznan University of Economics and Business  
**Bert George**  
Ghent University

# The Effect of the EU-Brand on Citizens' Trust in Policies: Research Article Replicating an Experiment

**Abstract:** *Karens et al. (2016) conducted an experiment to measure the effect of the European Union (EU) brand on citizens' trust in policies. Experiments conducted with economics students in Belgium, Poland, and The Netherlands showed a consistently positive and significant effect of applying the EU brand, on trust in the policies. This study presents seven replications conducted several years after Karens et al.'s experiments. The replications show no significant effects of the EU brand on trust in policies. These findings demonstrate that brand effects may vary over time. To identify a population effect size across all experiments, a meta-analysis was conducted. The meta-analysis shows that—overall—the EU-brand has a small but significant positive effect on citizens' trust in policies. The article tests earlier findings, and discusses intricacies of conducting replications. It elaborates explanations for the results in the replications, and the replication problems with experiments based on evaluative conditioning.*

## Evidence for Practice

- Brands of public organizations have the ability to influence citizens' perceptions of the organizations' policies.
- The effect of a public brand may vary over time, and even become non-significant during certain periods. Therefore, it is important to measure the impact of your brand longitudinally, instead of only once in time.

## The Importance of Replications in Public Administration

Public administration research has seen an upsurge of experimental research (e.g., Grimmelikhuijsen et al. 2017; James, Jilke, and van Ryzin 2017). This has helped the discipline to better study a wide range of causal effects, e.g., causal effects of public brands (Alon-Barkat and Gilad 2017; Teodoro and An 2017), red tape (Wesley and Tummers 2017), and cultural stereotypes in applicant-assessment (Baekgaard and George 2018).

### The Importance of Replication

However, following concerns in the wider social sciences about the robustness and replicability of findings in experimental research (Braver, Thoemmes, and Rosenthal 2014; Open Science Collaboration 2015), concerns about replicability of experiments have been raised in PA as well (Walker, James, and Brewer 2017). Context and timing are important in the field of public administration, which may create extra limitations in terms of replicability of experiments in PA research (Walker, James, and Brewer 2017). This raises questions about validity of earlier experimental research in PA, and calls for replication studies (Pedersen and Stritch 2018).

Replications are important for checking the validity of previous studies, and establishing the generalizability of findings to different populations and contexts (Klein et al. 2014; Pedersen and Stritch 2018; Schmidt 2009; Walker, James, and Brewer 2017). Nosek and Lakens (2014) explain that replications:

1. add data to increase precision of the effect size estimate via meta-analysis, and help to identify false positive research findings;
2. help to establish generalizability of effects, i.e., establish whether the sample, setting, and procedures that are assumed irrelevant in an experiment, are irrelevant indeed;
3. can help to identify boundary conditions for effects if they produce negative results.

This article presents the replication of an experimental research through evaluative conditioning with public brands. The original research was conducted by Karens et al. (2016). Evaluative conditioning pertains to changes in the evaluation of a stimulus (a policy in our case) as a result of merely pairing it with another positively or negatively perceived stimulus (the EU brand in our case) (Martin and Levey 1978; see also Alon-Barkat and Gilad 2017).

**Jasper Eshuis** is an associate professor at the Department of Public Administration and Sociology, Erasmus University Rotterdam. His research focuses on governance and management of emotions, public branding, and stakeholder participation. He has published widely, in journals such as *Public Administration Review*, *Public Administration* and *Public Management Review*. He is on the editorial board of *Public Administration Review*.

**Email:** eshuis@essb.eur.nl

**Thijs van de Geest** is an analyst at the Rabobank. He obtained his Master degree at the Department of Public Administration and Sociology, Erasmus University Rotterdam. His interest is on the cutting edge between public administration and business administration.

**Email:** thijsvandegeest@gmail.com

**Erik Hans Klijn** is a professor of public administration in the Department of Public Administration and Sociology, Erasmus University Rotterdam, the Netherlands. His research focuses on complex decisionmaking in networks, network management, public-private partnerships, branding, and media influence on governance.

**Email:** klijn@essb.eur.nl

**Joris Voets** is an associate professor of public management in the Department of Public Governance, Management and Finance at Ghent University, Belgium. His current research includes governance and performance of interorganizational networks and public-private partnerships, as well as service delivery by and government capacity of local governments. His recent work has been published in *Public Administration*, *Public Management Review*, and *Public Money and Management*.

**Email:** joris.voets@ugent.be

**Magdalena Florek** is an Associate Professor at the University of Economics and Business in Poznań, Poland. She is board member of the International Place Branding Association and Senior Fellow at the Institute of Place Management. Magdalena has consulted and coauthored numerous branding and promotion strategies for towns, cities and regions in Poland. She is Member of the Editorial Boards of, i.a., *Place Branding and Public Diplomacy* and *Journal of Destination Marketing and Management*.

**Email:** magdalena.florek@ue.poznan.pl

**Bert George** is a professor at the Department of Public Governance and Management at Ghent University, Belgium. His research focuses on strategy, behavior, and performance in public organizations using experimental, observational, and metaanalytical research methods.

**Email:** bert.george@ugent.be

*Public Administration Review*, Vol. 00, Iss. 00, pp. 1–11. © 2021 The Authors. *Public Administration Review* published by Wiley Periodicals LLC on behalf of American Society for Public Administration. DOI: 10.1111/puar.13367.

## ***This Article: Replicating an Experiment on the Effect of Branding on Policy Trust***

Karens et al. (2016) studied the effect of the European Union brand on citizen trust in policies. They tested whether the evaluation of two policies changed by pairing the policy with the EU-brand, building on the theoretical idea that “promotional symbols can change people’s attitudes through associative transfer of affect, or evaluative conditioning” (Alon-Barkat and Gilad 2017, 2). Karens et al. (2016) conducted experiments at lectures of economics students in Belgium, Poland, and The Netherlands. They found a consistently significant and positive effect of applying the EU brand, on trust in the policies.

This article researches whether those findings can be replicated in the same countries at the same series of lectures in 2017 and 2018, i.e., with the same population but a different sample and at a later moment in time.<sup>1</sup> It also studies whether the findings are replicable among other populations of students. As will be elaborated, some replications also included a conceptual improvement or additions. In terms of Walker, James, and Brewer (2017) this article thus presents exact replications, conceptual replications, and empirical generalizations. In terms of Pedersen and Stritch (2018), the value of this article is in addressing an empirically relevant issue (public branding), with low numbers of replications already existing, adding to internal validity through a conceptual improving (adding a fake brand); and testing the external validity by doing the study in new populations and new time periods. We also address recent calls (Braver, Thoemmes, and Rosenthal 2014; Open Science Collaboration 2015) for doing a meta-analysis on experiments and their replications to identify a population effect size across all studies.

This article is structured as follows: The next section discusses the research methods applied in the original studies and the replications. Subsequently, the findings of the individual replications and the meta-analysis are presented. In the fourth section, possible explanations for the findings are discussed. Finally, main conclusions and reflections on future research are formulated.

## **The Experiment: Measuring the Influence of the EU Brand on Trust in Policy**

The original research was a typical evaluative conditioning research whereby subjects’ attitudes toward a product/service (in this case a policy) were conditioned with a brand-type stimulus (in this case a logo and the brand name). This is a classic and common type of research in marketing and consumer sciences (e.g., Stuart, Shimp, and Engle 1987), and it has also been applied in Public Administration (Alon-Barkat and Gilad 2017; Teodoro and An 2017).

### ***Experimental Design of the Original Study***

The test units in the original study were academic economics students in the Netherlands, Belgium, and Poland. Experiments were conducted in lecture rooms at universities, in the period between April 2013 and June 2013. The students were randomly assigned by the primary researcher, to either the test group or the control group. Because no pre-measurement was conducted, this resulted in a post-test-only control group design (Morton and Williams 2010).

The dependent variables, trust in two policies, were measured in

both the control group and the test group by measuring trust in two policy proposals. The test group received two short descriptions of policy proposals to which EU brand elements (EU flag, name, and logo of the European Commission) were added, whereas the control group received the same policy proposals without any brand element (see Appendix in the Supporting Information that shows the full survey used for the treatment and control groups). The policy proposals in the experiment are based on actual interventions proposed by the European Commission policy plans but rewritten in such a way that they could be presented in a short text. Because it might be argued that different types of policy instruments are perceived differently by citizens, the two policy proposals are based on opposite types of policy instruments (Peters 2000): the first policy on digital skills was an example of an enabling policy instrument, whereas the second policy on air quality represents a coercive policy instrument. To prevent language problems, the experiment was stated in the mother tongue of the students: Dutch for the Netherlands and Belgium, and Polish for Poland.<sup>2</sup>

### ***Experimental Design of the Replications Study***

We conducted three forms of replication: direct replication, generalization, and conceptual replication. In total, we replicated the experiment seven times (for an overview see Table 1).

***Direct Replication.*** The primary aim of the replications was to conduct direct replications of the original study, so we kept the experimental procedures as similar as possible to the original studies, both during data gathering and data analyses. The experiments were replicated six times in 2017 and one time in 2018 (R1—R7, see Table 1).

***Generalization.*** Three replications of the seven were conducted in 2017 in the same economics courses as in 2013, only with new cohorts of students (R1-BE2017, R2-PL2017, R3-NL2017). The other three replications in 2017 took place in other courses in order to test the replicability of the experiments among other groups of students from other disciplines; i.e., to test the empirical generalization of the original studies (R5-NL2017, R6-NL2017, R7-PL2017). The latter three replications were in an Economics course with international students in Poland, and in two medical science courses in the Netherlands. In 2018, another replication (R4-BE2018) was conducted in the same course in Belgium to replicate the study in yet another year and research the influence of time.

***Conceptual Replication.*** In all replications of 2017, we added one measure to test the co-varying effect of ‘brand awareness’. In two replications, we added some new stimuli and/or measured extra dependent variables in order to extend the original research. R3 studied the moderating effect of the country of origin by adding either a German flag or a Cyprian flag and explaining that the EU policy was originally developed by either Germany or Cyprus. We thus tested whether a country of origin effect (e.g., Verlegh and Steenkamp 1999) influenced policy trust. In one study, we also assessed an ‘expert effect’ by adding a stimulus saying that the policy had been reviewed and found effective by an expert. We also tested a fake brand. This in response to criticism voiced in the literature that Karens et al. had not measured the possible effect of a fake brand (Alon-Barkat and Gilad 2017), and that Karens et al. therefore were unable to establish whether the effect of the EU

**Table 1** Overview of the Original Study and the Replications (BE = Belgium, PL = Poland, NL = Netherlands)

	O1-BE2013	O2-PL2013	O3-NL 2013	R1-BE2017	R2-PL2017	R3-NL2017	R4-BE2018	R5-NL2017 Medical Students (I)	R6-NL2017 Medical Students (II)	R7-PL 2017 International Students
Number of participants	66	89	58	111	137	95	149	88	115	84
Type of replication	Original study	Original study	Original study	Exact replication + Extension (DV)	Exact replication + Extension (DV)	Exact replication + Extension (extra stimulus)	Exact replication + Extension (DV)	Exact replication	Exact replication + extension (extra stimulus)	Exact Replication
Researchers conducting the experiment	Karens/Voets/	Karens/Florek	Karens/Welten	Meeuwissen/Voets	Meeuwissen/Florek	Meeuwissen/Welten	Van de Geest/Voets	Van de Geest/Kleinserink	Van de Geest/Goudzwaard	Van de Geest/Florek
Population and subjects ( $\sigma$ )	Economics students, Ghent University	Economics and Management students, Poznan University of Economics and Business	Economics students, Erasmus University Rotterdam	Same population different subjects	Same population different subjects	Same population different subjects	Same population different subjects	Different population different subjects	Different population different subjects	Different population different subjects
Stimuli	No brand/EU logo and EU name	No brand/EU logo and EU name	No brand/EU logo and EU name	No brand/EU logo and EU name	No brand/EU logo and EU name	No brand/EU logo and EU name/ German flag and German name	No brand/EU logo and EU name	No brand/EU logo and EU name	No brand/EU logo and EU name/ EU logo, EU name and policy expert	No brand/EU logo and EU name
Dependent variables	Trust in policy	Trust in policy	Trust in policy	Trust in policy/ Emotion/	Trust in policy/ Emotion/	Trust in policy/ Emotion/	Trust in policy/ Policy acceptance	Trust in policy/ Emotion/	Trust in policy/ Emotion/	Trust in policy/ Emotion/
Moderating variables				Brand awareness	Brand awareness	Brand awareness/ country of origin effect		Brand awareness/ country of origin effect	Brand awareness/ Expert effect	Brand awareness

brand would have been caused by *any* brand, rather than only by the EU brand.

As will be elaborated further on, no significant effects were found of any additional dependent variable or moderating variable.

### Experimental Procedure

The same questionnaire was used in the replications, and also students were randomly assigned by the primary researcher. The full questionnaires can be found in the Appendix in the Supporting Information. Slight differences with the original study may have occurred, because the researchers conducting the replications were different persons than with the original study in 2013. They had discussed with the original researcher what procedures he had followed in the lecture rooms during the experiments, particularly what instructions he had given, but slight differences in the exact wording and intonation may have occurred since there were no recorded instructions. Also, later we found out that the randomization in the class had not been exactly the same during the replications in 2017. During the original research, the questionnaire was handed out row by row; row A would get a questionnaire with stimulus, row B without, row C with stimulus, et cetera. During the replications in 2017, the questionnaires were handed out in a manner that each stimulus was handed out in equal numbers. This meant that each participant had a form differing from his or her neighbors (see Figure 1).

In 2018, another replication was conducted in the same course in Belgium (R4-BE2018) not only to study the influence of time, but also to conduct the randomization in the exact same way as in 2013.

### Research Findings: Results of the Replications

This section presents the descriptive statistics and the ANOVA analyses of both the original research (Karens et al. 2016) and the multiple replications. The ANOVA analyses are one-tailed (F-distribution is asymmetrical), one-way (we only focus on one categorical variable, namely, control versus treatment groups). As an inferential rule, we argue that an effect is significant when the  $p$ -value is smaller than .05. However, we report full  $p$ -values and confidence intervals for transparency and interpretation purposes. Moreover, for transparency and completeness purposes we also report the results of the Levene's test and Kruskal-Wallis test.

The descriptive statistics (see Table 2) show that trust in the EU went up over the years. Further, the average scores on trust in both policies are slightly higher in the replications than in the original experiments.

### Analyzing the Replications

Results of the replications (see Tables 3–5) show that no significant effects were found based on our inferential rule of  $p < .05$ .

Importantly, we do find that some experiments obtained ANOVA

Original experiment and replication 2018

Replications 2017

Row A	s s s s s s s s s s s s s s s s	s c s c s c s c s c s c s c s c
Row B	c c c c c c c c c c c c c c c c	c s c s c s c s c s c s c s c s
Row C	s s s s s s s s s s s s s s s s	s c s c s c s c s c s c s c s c
Row D	c c c c c c c c c c c c c c c c	c s c s c s c s c s c s c s c s

Figure 1 Randomization in the Original Experiment and Replications (s = Stimulus Group, c = Control Group)

Table 2 Descriptive Statistics of Variables Used in the Original Studies (2013) and Replications (2017 and 2018)

Variable	01-BE 2013	02-PL 2013	03-NL2013	R1-BE2017	R2-PL 2017	R3-NL2017	R4-BE 2018	R5-NL2017 Medical Students (I)	R6-NL2017 Medical Students (II)	R7-PL2017 International Students
Number of participants	66	89	58	111	137	95	149	88	115	84
Test/control	33/33	43/46	29/29	37/37/37*	68/69	32/32/31**	68/81	31/30/27***	38/46/31****	43/41
Man/Woman	29/37	39/50	40/18	52/59	46/88	57/38	71/78	36/52	24/91	41/43
Age (σ)	20.36 (1.432)	21.70 (1.95)	21.05 (1.76)	21.39 (3.82)	20.07 (0.93)	20.14 (1.01)	22.24 (6.70)	20.09 (1.49)	19.10 (1.05)	22.10 (1.96)
Trust in digital skills policy (σ)	6.40 (1.08)	6.03 (1.47)	5.63 (1.22)	6.31 (1.06)	6.56 (1.45)	6.22 (1.25)	6.40 (1.18)	6.11 (1.23)	6.31 (1.06)	5.97 (1.58)
Trust in air quality policy (σ)	6.73 (1.27)	6.23 (1.37)	6.15 (1.10)	6.8.(1.49)	6.79 (1.62)	6.93 (1.09)	6.90 (1.22)	6.92 (1.17)	7.32 (0.96)	6.55 (1.80)
Trust in European Union	6.86 (1.64)	5.35 (2.02)	5.97 (1.68)	7.07 (1.37)	5.42 (2.15)	6.24 (1.62)	6.79 (1.60)	6.42 (1.52)	6.86 (1.40)	4.80 (2.22)

\* Fake brand.  
 \*\* German stimulus.  
 \*\*\* Cyprus stimulus.  
 \*\*\*\* Expert effect.

Table 3 Belgium Experiments 2013, 2017, 2018: Results of the ANOVA and Kruskal-Wallis Tests Among Participants, with Experimental Treatment as Independent Variable and Trust in two EU Policies as Dependent Variables

	01-BE 2013 Digital Skills Policy 2013	01-BE 2013 Air Quality Policy 2013	R1-BE2017 Digital Skills Policy 2017	R1-BE2017 Air Quality Policy 2017	R4-BE 2018 Digital Skills Policy 2018	R4-BE 2018 Air Quality Policy 2018
Mean control group (σ)	5.98 (0.987) [5.63–6.33]	6.31 (1.058) [5.93–6.68]	6.27 (1.19) [5.88–6.67]	6.52 (1.68) [5.96–7.08]	6.48 (1.07) [6.22–6.74]	6.86 (1.23) [6.56–7.15]
[95% Conf. Interval]						
Mean experimental group (σ)	6.83 (1.023) [6.47–7.19]	7.17 (1.334) [6.69–7.65]	6.36 (0.98) [6.03–6.69]	7.11 (1.32) [6.67–7.55]	6.33 (1.26) [6.05–6.61]	6.93 (1.21) [6.66–7.20]
[95% Conf. Interval]						
ANOVA (p-value)	11.764 (0.001)	8.389 (0.005)	0.116 (0.734)	2.819 (0.097)	0.587 (0.445)	0.126 (0.723)
Levene's test (p-value)	0.024 (0.88)	0.495 (0.48)	0.044 (0.84)	0.457 (0.50)	1.797 (0.18)	0.001 (0.98)
Kruskal-Wallis	10.811 (0.001)	8.171 (0.004)	0.001 (0.974)	2.231 (0.135)	0.367 (0.545)	0.097 (0.756)

values that had a *p*-value > .05 but < .10, which indicates that these were close to our cut-off point. However, because we decided upfront to use *p* < .05 as our inferential rule, these were still classified as non-significant. The effects were non-significant in the groups with economics students (similar to original research), nor in the groups with Dutch medical students (different discipline), nor among the Polish international students of economics (multiple home countries).

Since the original studies did not measure the possible effect of a fake brand, these studies did not establish whether the effect of the EU brand would also have been caused by any brand, rather than only by the EU brand. Therefore, one replication also tested the effect of a fake brand (see Table 6).

Table 6 shows that the subjects in the fake brand group on average have a slightly higher level of trust than the control group (no

brand at all), but lower than the experimental group (with the EU brand). However, the difference between the fake brand and the control group is insignificant based on our inferential rule of *p* < .05 (and not even close), and the same goes for the difference between the control group and the EU brand group. This means that the fake brand does not have a significant effect on policy trust. The difference between the control group and the experimental group is also insignificant, implying that in these replications of the experiment the EU brand did not have a significant effect either.

**Trust in the EU as Explanation for Non-significant Replications?**

In 2013, the EU brand had a significant effect on policy trust. Thus in 2013, the EU brand did create certain positive associations with the branded policies, which resulted in a

**Table 4** Poland Experiments 2013 and 2017: Results of the ANOVA and Kruskal-Wallis Tests Among Participants, with Experimental Treatment as Independent Variable and Trust in two EU Policies as Dependent Variables

	02-PL 2013 Digital Skills Policy 2013	02-PL 2013 Air Quality Policy 2013	R2-PL 2017 Digital Skills Policy 2017	R2-PL 2017 Air Quality Policy 2017	R7-PL2017 Digital Skills Policy 2017 International Students	R7-PL2017 Air Quality Policy 2017 International Students
Mean control group ( $\sigma$ )	5.49 (1.501)	5.77 (1.230)	6.41 (1.50)	6.71 (1.82)	5.88 (1.44)	6.52 (1.68)
[95% Conf. Interval]	[5.04–5.93]	[5.40–6.13]	[6.05–6.77]	[6.27–7.15]	[5.41–6.34]	[5.97–7.07]
Mean experimental group ( $\sigma$ )	6.61 (1.191)	6.75 (1.327)	6.71 (1.50)	6.88 (1.39)	6.06 (1.71)	6.57 (1.92)
[95% Conf. Interval]	[6.25–6.98]	[6.34–7.16]	[6.37–7.05]	[6.53–7.22]	[5.53–5.59]	[5.97–7.17]
ANOVA ( $p$ -value)	15.253 (0.000)	13.077 (0.001)	1.501 (0.223)	0.369 (0.544)	0.283 (0.596)	0.013 (0.911)
Levene's test ( $p$ -value)	2.695 (0.10)	0.007 (0.93)	0.451 (0.50)	1.480 (0.23)	0.745 (0.39)	0.716 (0.40)
Kruskal-Wallis ( $p$ -value)	13.746 (0.000)	13.649 (0.000)	1.860 (0.173)	0.009 (0.924)	0.247 (0.619)	0.107 (0.743)

**Table 5** Netherlands Experiments 2013 and 2017: Results of the ANOVA and Kruskal-Wallis Tests among Participants, with Experimental Treatment as Independent Variable and Trust in Two EU Policies as Dependent Variables

	03-NL2013 Digital Skills Policy 2013	03-NL2013 Air Quality Policy 2013	R3-NL2017 Digital Skills Policy 2017	R3-NL2017 Air Quality Policy 2017	R5-NL2017 Digital Skills Policy 2017 Medical Science Students (I)	R5-NL2017 Air Quality Policy 2017 Medical Science Students (I)	R6-NL2017 Digital Skills Policy 2017 Medical Science (II)	R6-NL2017 Air Quality Policy 2017 Medical Science (II)
Mean control group ( $\sigma$ )	5.26 (1.125)	5.90 (1.140)	6.29 (1.12)	7.14 (0.85)	5.97 (1.31)	6.62 (1.16)	6.30 (0.93)	7.36 (0.83)
[95% Conf. Interval]	[4.84–5.69]	[5.46–6.33]	[5.87–6.71]	[6.81–7.45]	[5.46–6.48]	[6.17–7.07]	[6.01–6.59]	[7.09–7.64]
Mean experimental group ( $\sigma$ )	6.00 (1.210)	6.41 (1.012)	6.16 (1.38)	6.74 (1.26)	6.24 (1.17)	7.23 (1.12)	6.25 (0.95)	7.12 (1.04)
[95% Conf. Interval]	[5.54–6.46]	[6.02–6.79]	[5.67–6.66]	[6.29–7.20]	[5.79–6.68]	[6.79–7.66]	[5.92–6.58]	[6.75–7.49]
ANOVA ( $p$ -value)	5.750 (0.020)	3.263 (0.076)	0.166 (0.685)	1.936 (0.169)	0.656 (0.421)	3.998 (0.051)	0.054 (0.817)	1.194 (0.278)
Levene's test ( $p$ -value)	0.144 (0.71)	0.334 (0.57)	0.405 (0.53)	2.761 (0.10)	0.158 (0.69)	0.350 (0.56)	0.035 (0.85)	1.034 (0.31)
Kruskal-Wallis ( $p$ -value)	7.185 (0.007)	5.363 (0.021)	0.097 (0.756)	1.458 (0.227)	0.921 (0.337)	4.175 (0.041)	0.001 (0.975)	0.937 (0.333)

**Table 6** Effect of Fake Brand Compared with EU Brand and Control Group

	Digital Skills Policy 2017	Air Quality Policy 2017
Mean control group ( $\sigma$ )	6.27 (1.19)	6.52 (1.68)
[95% Conf. Interval]	[5.88–6.67]	[5.96–7.08]
Mean experimental group ( $\sigma$ )	6.35 (0.98)	7.11 (1.42)
[95% Conf. Interval]	[6.03–6.69]	[6.67–7.55]
Mean fake brand Group	6.29 (1.02)	6.85 (1.49)
[95% Conf. Interval]	[5.95–6.63]	[6.37–7.33]
ANOVA ( $p$ -value)	0.067 (0.935)	1.470 (0.234)
Levene's test ( $p$ -value)	0.261 (0.77)	0.256 (0.78)
Kruskal-Wallis ( $p$ -value)	0.033 (0.98)	2.394 (0.30)

significant increase of trust in those policies. In theory, the reason why we did not find significant effects of the EU brand in 2017 and 2018 might be that the EU brand was perceived less positively in 2017 and 2018, and as a result the EU brand may have transferred less positive associations to the policies, which then did not significantly increase trust in those policies. This does not necessarily mean that the EU brand does not transfer any associations, it only means that the associations do not significantly change the evaluation of the policies. In other words, when the respondents in 2017 and 2018 perceived the EU less positively than in 2013, it is possible that the brand did transfer associations from the brand to the policy, only those associations were less positive and therefore they did not significantly increase trust in the policies as they did in 2013.

**Table 7** Meta-Analytical Findings

	Number of Effect Sizes	Population Effect Size (SMD)	95% Confidence Interval	z-Score	$p$ -Value
Digital skills policy	10	.249	[.027, .472]	2.20	.028
Air quality policy	10	.232	[.004, .461]	1.99	.047

To test whether this could explain our non-significant results, we measured the level of trust in the EU (see also Table 1), and tested whether trust in the EU moderated the effect of the brand on policy trust. The reasoning behind this analysis was: if trust in the EU went down from 2013 to 2017 and 2018, this may explain why the brand did not have a significant positive effect in 2017 and 2018. Table 1 shows that trust in the EU went up, instead of going down. An additional regression analysis to check whether trust in the EU moderated the effect of the brand, also showed that trust in the EU did not significantly moderate the effect of the EU brand. Thus, the lack of a significant positive effect in the EU brand in the replications, cannot be attributed to less trust in the EU. Interestingly, this indicates that the effect of a public organizational brand through evaluative conditioning is not necessarily determined by trust in that organization. The evaluative conditioning effects of public brands are more complex, as we will

discuss more elaborately further on. Thereby, it must be noted that we measured only *trust* in the EU, which is only one dimension of the possible perceptions regarding the EU. We cannot completely rule out that other perceptions of the EU became more negative and prevented a significant positive impact of the EU brand on trust in the EU policy. However, along with many political scientists studying the EU (e.g., Leconte 2010; Munoz, Torcal, and Bonet 2011) we do think that trust in the EU is a good overall indicator of how people perceive the EU. Also, it is logical to expect that specifically trust in the EU would significantly affect trust in EU-branded policies.

### Studying the Whole Picture: A Meta-Analysis of all Experiments Together

Analyzing whether replications yield significant results or not, and how their effect sizes compare to the original study is an important but limited step in replication studies (see e.g., Klein et al. 2014; Open Science Collaboration 2015). It is also important to analyze what the effect of the EU-brand is when looking at all data combined. Braver, Thoemmes, and Rosenthal (2014) underline the importance of conducting meta-analyses in order to get a robust insight into the significance and strength of an effect. Braver et al. (2014, 334) argue that “the resulting pooled estimate generally is more trustworthy because it is based on far more data than each individual study”. Their article shows that even when replications of an experiment do not yield significant results, the effect of the stimulus in the overall dataset can be more significant than in the original studies alone due the higher N. In support of this line of reasoning, the Open Science Collaboration (2015) found that while only 36% of the replicated experiments yielded significant findings, 68% of the studies had significant results when looking at the overall dataset that combined the

original study and the replications. Recent meta-analyses in public administration have typically focused on integrating studies on a subject such as strategic planning (George, Walker, and Monster 2019), red tape (George et al. 2020) or performance management (Gerrish 2016). The meta-analysis conducted within this study has two advantages compared to these other meta-analyses in our field: (1) there is little to no heterogeneity in terms of treatment (EU brand), dependent variables (trust in policies), and type of population (economics students) in our experiments and (2) all studies included in the meta-analysis are experimental and allow causal inference. As such, the identified population effect size across all experiments cannot be distorted due to endogeneity in individual studies and remains comparable to the interpretation of the effect of each individual experiment. A random-effects meta-analysis is conducted based on the means of the control and treatment groups, their sample sizes and their standard deviations as reported in the original study and its replications. Two analyses are conducted—one focused on trust in a digital skills policy, and the other focused on trust in the air quality policy. Table 7 presents the results from both analyses.

The results from Table 7 suggest that across all studies, the effect of the EU-brand on trust in policies is positive and significant. However, when we look at the SMD-value, we can—based on Cohen (1988)—conclude that the strength of this effect is small (SMD = .2). Figures 2 and 3 show the SMD of each individual study, and the weight it carries in the overall population effect size.

### Discussion: How Can the Different Results of the Replications Be Explained?

This section discusses possible explanations for the findings of the original study being confirmed in the overall meta-analysis,

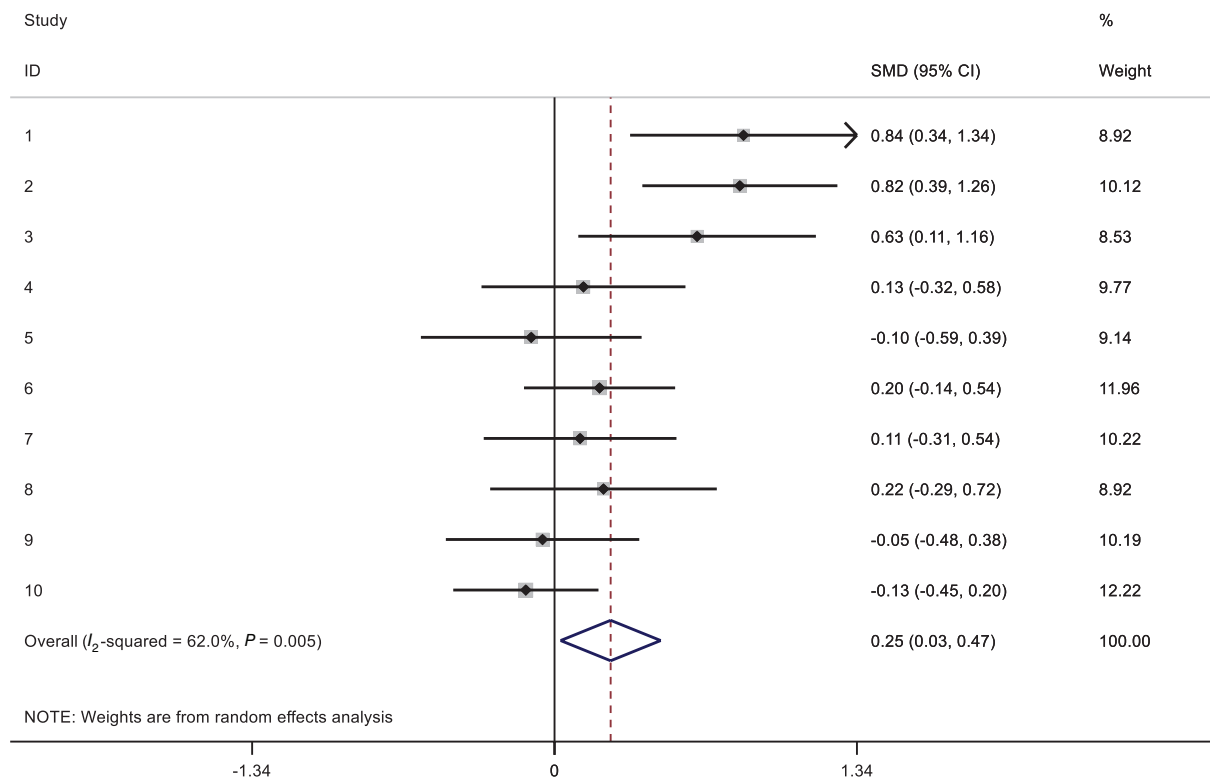
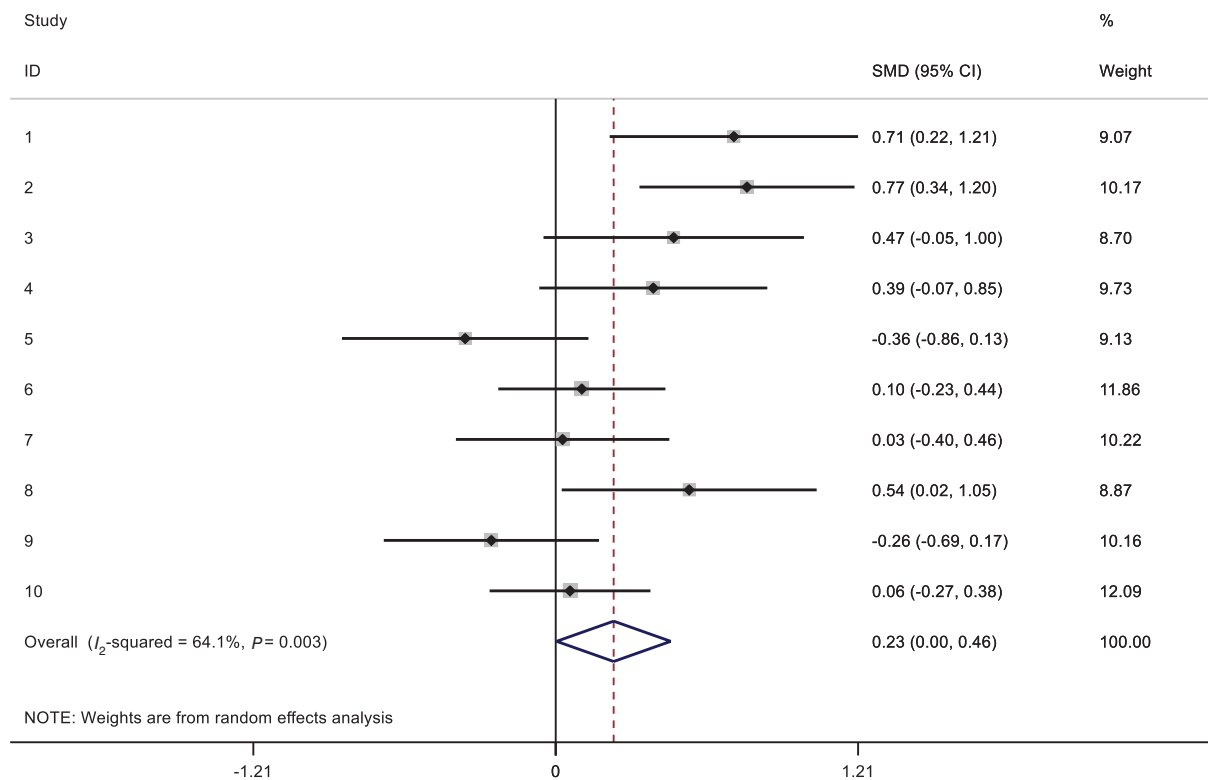


Figure 2 Studies’ SMD and Their Weight in Overall SMD for the Digital Skills Policy



**Figure 3 Studies' SMD and Their Weight in Overall SMD for Air Quality Policy**

but not being replicated in the replications. It discusses whether different findings in the replications can be attributed to procedural variations, setting and context, or the type of experiment.

**Possible Causes for the Difference between Original Experiments and Replications**

One reason why the overall meta-analysis shows a significant (but small) positive effect of the EU-brand, is of course the significant and strong positive effects of the brand in the original studies in 2013. This effect was so strong that there is an overall significant effect. Thereby, also most replications show some positive effect of the brand, and although these are not significant, they do contribute to the overall significant and positive effect size.

The non-significant results in the replications may be caused by slight differences in the set-up between the original study and the replications, even in studies such as this one where exact replication was attempted. We discuss some differences between the original study and the replications.

- *Differences in set up of the experiments;* the replications in 2017 differed slightly from the original studies in how the control groups and stimulus groups were divided over the lecture room (see Figure 1). During the replications, students may have witnessed that their neighbors had a survey with one different element, namely, an EU logo on it. One student (who was in a control group) explicitly mentioned during a focus group session afterward, that he had seen the EU logo from his neighbor. Although we had several focus groups and chats after the experiments and we received no further signals about this, it cannot be ruled out that for some subjects the experiment was thus compromised. To test for this possibility, we conducted

the experiment once again in Belgium in 2018, using the exact same way of dividing the groups as Karens did, and also then we could not replicate the findings of the original study. This does suggest that it was not the different randomization which caused the non-significant results in 2017, but of course the experiment in 2018 provides only indirect evidence.

- *Differences in researchers conducting the experiment;* The individuals conducting the replications in the lecture-rooms were partly the same individuals (the lecturer was the same) but there were also new individuals (Karens was replaced by ANONIMIZED). Notwithstanding close communication with Karens about the procedures of the experiments, the new individuals may have triggered slightly different associations. In principle, we do not expect a large bias of this, because both the control group and the stimulus groups were confronted with the new individuals, so this bias should be prevented by the randomization. However, in theory it is possible that the new researchers triggered particular associations that become particularly salient in connection with the EU brand, thus creating a non-random effect (a differing effect for the stimulus group than for the control group).
- *Differences in research population;* Even though the population was the same in the original study and the replications (economic students in three specific courses) the sample in the replication was different in the sense of being new cohorts of students. The fact that the samples are drawn from the same populations of economics students in 2013 and 2017/2018 ensured that not only their education is the same, but also their age and gender division (there are no significant differences in the age or gender division between the original study and the replications). Nevertheless, it is possible that the new cohorts had different associations with the EU brand.

Although trust in the EU did not change very much between the cohorts, it is possible that the associations triggered by the EU brand did change. In hindsight, it would have been relevant to have included questions in the survey about the subjects' associations with the EU brand. This underlines the importance of control variables and the inclusions of variables that may help to further explain the effects of a stimulus.

Reflecting on the set-up of the original study and the replications, it seems unlikely that the differences caused the non-replication because the differences are small. Also, the 2018 replication which was highly similar to the original studies, did not produce the same results as the original studies.

Also, differences in sampling and experimental setting were not large. Most replications were drawn from the same population, and main demographic characteristics of the samples were very similar. Thus, this is not likely to explain the deviant results. This idea is further supported by the finding from both the Open Science Collaboration (2015) and the Many labs study (Klein et al. 2014) that the original effect size is the most important factor to predict replication success, not sampling or the setting used.

Another possibility is that there may have been changes in the wider context between the original study and the replication. Is it possible that changes over time in the societal or political context made people look differently at the EU brand, which in turn then had different (less strong and less significant) effects on policy trust? e.g., a major and highly visible process such as the Brexit may lead to different associations with the EU brand. Is it possible that a stimulus such as the EU brand has different effects in different contexts, because the associations that the stimulus triggers vary with the context? This brings us to a discussion of the replicability of evaluative conditioning experiments.

### ***Some Reflections on Evaluative Conditioning***

Existing literature suggests that the results of evaluative conditioning (EC) are often fragile, instead of robust (Jones, Olson, and Fazio 2010; Rozin, Wrzesniewski, and Byrnes 1998). On the basis of 20 studies, Rozin, Wrzesniewski, and Byrnes (1998) argue that all together there is a significant effect of evaluative conditioning, but that the significant effects are (inexplicably) absent in certain groups of studies. Hofmann et al. (2010) underline the existence of many conflicting findings in the literature.

Notwithstanding the fact that our meta-analysis confirms the overall significant and positive effect of the EU-brand stimulus on policy trust, this study affirms the variability of findings in evaluative conditioning experiments in the field of Public Administration; the original studies found strong significant effects while the replications did not. In the following, we elaborate on how evaluative conditioning works, in order to get better insight in (varying) outcomes of evaluative conditioning experiments.

Evaluative conditioning tries to add new associations to the network of associations in people's minds. In this study, the researchers tried to add new associations (triggered by the EU-brand) to the existing associations about two public policies (air quality and digital skills). The evaluations of the policies were

significantly influenced by the brand during the original studies, but not during the replications.

Earlier research on evaluative conditioning experiments can provide possible explanations. Simply put, it could be that existing associations about the policies were so strong during the times of the replications in 2017 and 2018, that they were not changed by the associations triggered by the EU brand. This can be explained as follows. Literature on evaluative conditioning has brought forward that with attitude objects that are familiar (e.g., a racial group or, in this case two policies) it is unlikely that evaluative conditioning would significantly affect explicit attitudes (Gibson, Lodge, and Woodson 2014) such as trust. Following this line of reasoning, it could thus be that the attitude object used in the experiment, i.e., public policies on air quality and digital skills, had become so familiar to the research subjects (students of economics) in 2017 and 2018, that they already had an extensive associative network regarding the policies. Because the associative network was extensive, it was not significantly altered by a short and one-time exposure of the EU-brand. This short exposure could not add sufficient new connections to the already existing associative network regarding the two policies to significantly alter the trust in the policies (compare Gibson 2008; Gibson, Lodge, and Woodson 2014). This could be an explanation why the EU-brand did alter associations in 2013, but not in 2017 and 2018 because in 2017 and 2018 air quality policies and digital skills policies by the EU had existed for a longer time and thus it is possible that they were more familiar to the groups from 2017 and 2018 than to the groups from 2013. Environmental policies were likely more 'present' because of the global warming discussion which has received a lot of attention. But, discussions about internet and digital policy have received much more attention in the media, for instance after the US elections of 2016 and the discussions about fake news spread through social media.

### **Conclusion**

On the basis of the overall meta-analytic evidence across all studies conducted, we conclude that the effect of the EU-brand on policy trust is significant and positive, albeit rather small in terms of power. However, it is important to note that the significant and positive effect of the EU brand on policy trust found by Karens et al. (2016) could not be replicated in the seven replications that were conducted in three countries, among both the same population of students and other populations of students in two different years (2017 and 2018). From the difference in findings between the experiments in 2013 compared with the experiments in 2017 and 2018, we conclude that the significant and positive effect of the EU brand on policy trust as reported in Karens et al. (2016) is not constant throughout the years, and may be different in various time periods. In other words, effects of public brands are dynamic, and they can vary over time.

This study has discussed several possible explanations for the results in the replications. Given the fact that the studies were replicated in very similar ways to the original experiments, and that the small changes that we could detect in the experimental setup of the replications most likely have equal effects in the control group and the experiment group (the effects would be thus neutralized



by the randomization), it is our contention that differences in the experimental setup are unlikely to explain the non-replication.

A more likely explanation in our view, is a change in the wider context between 2013 (when the original study was conducted), and 2017 and 2018 (when the replications were conducted). Those changes that took place over time may have caused different associations with the EU brand (the unconditioned stimulus) and with the policies (the conditioned stimulus) during the replications, which then in turn decreased the strength and significance of the positive effect of the brand on trust in EU policies. This shows a limitation with this type of experiments (evaluative conditioning experiments), namely, that the associations triggered by a certain conditioning stimulus may be highly context dependent, and that also the associations people already have with the unconditioned stimulus may differ over time.

The conclusion about varying brand effects caused by changes in associations that a brand triggers, is not only methodologically relevant but also invites theoretical elaboration. Theoretically, brands can be viewed as sets of associations about a certain object in the minds of people (Keller 1993). Public brands thus consist of nodes of information that form networks of associations in the minds of citizens. Various nodes can be “recalled from memory when a node is stimulated from rest by a process known as activation” (Smith and French 2009, 211). Literatures have explained that associative networks are dynamic because ongoing experience with the branded object may change brand associations, especially for immature or new brands (Kunkel et al. 2016). Our study suggests something additional, namely, that in the public realm also the associations that are activated by a *mature* brand can be dynamic. Because public brands are commonly subject of public debate wherein a rich variety of qualities may be ascribed to the brand (Eshuis and Klijn 2012), people may hold not only many perceptions of the brand but also ambiguous perceptions (see Gibson 2008). Thus, many differing associations can potentially be triggered by a mature public brand such as the EU brand. Also public attention may shift to different aspects of the brand over time. In case of the EU, public attention may shift, e.g., from a successful closure of a Brexit deal to failure in governing migration. Thus, with public brands people may hold many nodes of information, often mixed and even contradictory, and which node is activated most easily, depends on the situation in the public environment and citizens’ current attitudes. As Kunkel et al. (2016, 120) write, which associations are triggered depends on ‘the temporal context’ wherein the branded object is evaluated. This is in line with studies in Public Administration that elaborate how time may influence actor’s perceptions, and thus become a factor to explicitly consider in scientific research (e.g., Eshuis and van Buuren 2014; Pollitt 2008). Further research is required to measure which particular associations are triggered by public brands in different (temporal) contexts.

The conclusion that the overall effect of the EU-brand across all studies is significant even though only the original studies showed significant effects and none of the replications did, may raise the question whether the result of a meta-analysis would remain the same if more replications were carried out. The answer to that question is unknown at this stage. Particularly, the results of evaluative conditioning experiments vary across replications

(but this is also true for other experiments, e.g., in the field of social psychology as the famous replication by the Open Science Collaboration from 2015 demonstrated). Social scientific knowledge based on experiments is not definitive. The best that scientists can do is to conduct regular replications to test whether certain findings still hold, and publish the results of the new replications and the overall meta-analyses in order to provide knowledge based on the most complete and elaborate dataset as possible. Moreover, if enough replications become available, meta-analyses can also be expanded by including meta-regression analyses (see George et al. 2020; George, Walker, and Monster 2019) which can identify sources of variation in effect sizes overall and thus help to identify potential reasons as to why effect sizes might vary across replications.

### **Future Research**

What does this study imply for future experimental research in Public Administration? First of all, this study confirms the call for and the importance of meta-analysis to assess the significance and strength of an effect across original studies and replications thereof. As Braver, Thoemmes, and Rosenthal (2014) have argued, the pooled estimate in meta-analyses is based on more data than studies that use either the original study or only the replications.

But our findings also indicate that results of evaluative conditioning experiments in PA are contextual, and show significant limitations when it comes to generalizability across contexts. With the growing popularity of Behavioral Public Administration, and the increasing importance of experiments, these observations are more relevant for Public Administration than ever.

This also calls for future experimental research in which contextual variables are included, so that their effects can be assessed also during replications. Not just for evaluative conditioning experiments but all experiments in Public Administration, since most experimental design in Public Administration include elements of the real world and thus results may vary across contexts (Walker, James, and Brewer 2017).

Further, this study did bring out the empirical variance in brand effect, but it did not explain the underlying mechanism. So called ‘mechanism experiments’ (Ludwig, Kling, and Mullainathan 2011) could be conducted to test various mechanisms that possibly explain the (varying) effects of public brands. Mechanism experiments do not test whether a certain brand has a significant effect or not, but directly test the effect of certain causal mechanisms (see Ludwig, Kling, and Mullainathan 2011). Relevant future mechanism experiments could, e.g., study the mechanism of ‘making a policy official by adding a governmental brand’, by adding various official governmental brands and non-official brands to policies. Other mechanism experiments could test mechanisms based on the specific associations that the EU may trigger such as ‘expertise’, ‘power’, ‘stability’, ‘welfare’, ‘bureaucratic’, or ‘threat to autonomy’. Also the effects of various moderating mechanisms could be studied, such as emotional attachment to the brand, or contextual moderators such as characteristics of news coverage.

Most experiments in Public Administration differ from the very fundamental experiments often seen in economics or psychology, where, e.g., risk taking is reduced to some core features and

translated into a very basic game that bears no direct reference to actual risks in the real world (e.g., the Balloon Analogue Risk Task where participants pump balloons with different explosion points shown on screens (Lejuez et al. 2002)). Real-life associations are deliberately avoided in those experiments, in order to reduce risks of confounding factors, and find the fundamental mechanisms at play that are independent of a specific context. However, stripping experiments from contextual influences comes at the cost of low ecological validity, and the question then remains whether the effects found in the fundamental experiment hold in various real-world contexts. In our view, it is relevant for Public Administration as a discipline to conduct both fundamental experiments that establish fundamental mechanisms, but also experiments that test mechanisms in settings that resemble the real world.

## Notes

1. The authors thank Rene Karens for his cooperation with the replications, Robin Meeuwissen for implementing replications in the Netherlands and Poland, and Welten, Kleinsierink, and Goudzwaard for allowing the experiments during their lectures.
2. Measurement and scales. Scales measuring trust in the policy were theoretically based on work by Mayer, Davis, and David Schoorman (1995), Grimmelikhuijsen (2012) and McKnight, Choudhury, and Kacmar (2002). Closely following these literatures, Karens et al. (2016) used a scale of six items that measured trust in the benevolence, goodwill and capability of both policies. The reliability of the scale was high for both policies (Cronbach's alpha of 8.6 and 8.7).

## References

- Alon-Barkat, Saar, and Sharon Gilad. 2017. Compensating for Poor Performance with Promotional Symbols: Evidence from a Survey Experiment. *Journal of Public Administration Research and Theory* 27(4): 661–75.
- Baekgaard, Martin, and Bert George. 2018. Equal Access to the Top? Representative Bureaucracy and Politicians' Recruitment Preferences for Top Administrative Staff. *Journal of Public Administration Research and Theory* 28(4): 535–50. <https://doi.org/10.1093/jopart/muy038>.
- Braver, Sanford L., Felix J. Thoenemmes, and Robert Rosenthal. 2014. Continuously Cumulating Meta-Analysis and Replicability. *Perspectives on Psychological Science* 9(3): 333–42.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.
- Eshuis, Jasper, and Erik-Hans Klijn. 2012. *Branding in Governance and Public Management*. Abingdon: Routledge.
- Eshuis, Jasper, and Arwin van Buuren. 2014. Innovations in Water Governance: The Importance of Time. *International Review of Administrative Sciences* 80(2): 401–20.
- George, Bert, Richard M. Walker, and Joost Monster. 2019. Does Strategic Planning Improve Organizational Performance? A Meta-Analysis. *Public Administration Review* 79(6): 810–9.
- George, Bert, Sanjay K. Pandey, Bram Steijn, Adelen Decramer, and Mieke Audenaert. 2020. Red Tape, Organizational Performance and Employee Outcomes: Meta-Analysis, Meta-Regression and Research Agenda. *Public Administration Review* Early View.
- Gerrish. 2016. The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis. *Public Administration Review* 76(1): 48–66.
- Gibson, Bryan. 2008. Can Evaluative Conditioning Change Attitudes toward Mature Brands? New Evidence from the Implicit Association Test. *Journal of Consumer Research* 35: 178–88.
- Gibson, James L., Milton Lodge, and Benjamin Woodson. 2014. Losing, but Accepting: Legitimacy, Positivity Theory, and the Symbols of Judicial Authority. *Law and Society Review* 48: 837–66.
- Grimmelikhuijsen, Stephan G. 2012. Linking Transparency, Knowledge and Citizen Trust in Government: An Experiment. *International Review of Administrative Science* 78(1): 50–73.
- Grimmelikhuijsen, Stefan, Sebastian R. Jilke, Olsen, and L. Tummers. 2017. Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review* 77(1): 45–56.
- Hofmann, Wilhelm, Jan De Houwer, Marco Perugini, Frank Baeyens, and Geert Crombez. 2010. Evaluative Conditioning in Humans: A Meta-Analysis. *Psychological Bulletin* 136: 390–421. <https://doi.org/10.1037/a0018916>.
- James, Oliver, Sebastian R. Jilke, and Greg G. van Ryzin. 2017. *Experiments in Public Management Research: Challenges and Contributions*. Cambridge: Cambridge University Press.
- Jones, Christopher R., Michael A. Olson, and Russell H. Fazio. 2010. Evaluative Conditioning: The 'how' Question. *Advances in Experimental Social Psychology* 43: 205–55.
- Karens, Rene, Jasper Eshuis, Erik Hans Klijn, and Joris Voets. 2016. The Impact of Public Branding: An Experimental Study on the Effects of Branding Policy on Citizen Trust. *Public Administration Review* 76(3): 486–94.
- Kaufmann Wesley, and Lars Tummers. 2017. The Negative Effect of Red Tape on Procedural Satisfaction. *Public Management Review* 19(9): 1311–27.
- Keller, Kevin.L. 1993. Conceptualizing Customer-Based Brand Equity. *Journal of Marketing* 57(1): 1–22.
- Klein, Richard A., Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams, Jr., Štěpán Bahník, Michael J. Bernstein, Konrad Bocian, Mark J. Brandt, Beach Brooks, Claudia Chloe Brumbaugh, Zeynep Cemalcilar, Jesse Chandler, Winnee Cheong, William E. Davis, Thierry Devos, Matthew Eisner, Natalia Frankowska, David Furrow, Elisa Maria Galliani, Fred Hasselman, Joshua A. Hicks, James F. Hovermale, S. Jane Hunt, Jeffrey R. Huntsinger, Hans IJzerman, Melissa-Sue John, Jennifer A. Joy-Gaba, Heather Barry Kappes, Lacy E. Krueger, Jaime Kurtz, Carmel A. Levitan, Robyn K. Mallett, Wendy L. Morris, Anthony J. Nelson, Jason A. Nier, Grant Packard, Ronaldo Pilati, Abraham M. Rutchick, Kathleen Schmidt, Jeanine L. Skorinko, Robert Smith, Troy G. Steiner, Justin Storbeck, Lyn M. van Swol, Donna Thompson, A.E. van't Veer, Leigh Ann Vaughn, Marek Vranka, Aaron L. Wichman, Julie A. Woodzicka, and Brian A. Nosek. 2014. Investigating Variation in Replicability: A 'Many Labs' Replication Project. *Social Psychology* 45(3): 142–52.
- Kunkel, Thilio, Jason P. Doyle, Daniel C. Funk, Du James, and Health McDonald. 2016. The Development and Change of Brand Associations and their Influence on Team Loyalty over Time. *Journal of Sport Management* 30(2): 117–34.
- Leconte, Cecile. 2010. *Understanding Euro-scepticism*. Basingstoke: Palgrave Macmillan.
- Lejuez, C.W., J.P. Read, C.W. Kahler, J.B. Richards, S.E. Ramsey, G.L. Stuart, D.R. Strong, and R.A. Brown. 2002. Evaluation of a Behavioral Measure of Risk Taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied* 8: 75–84.
- Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2011. Mechanism Experiments and Policy Evaluations. *Journal of Economic Perspectives* 25(3): 17–38.
- Martin, I., and Levey A. B. 1978. Evaluative Conditioning. *Advances in Behaviour Research and Therapy* 1: 57–102.
- Mayer, Roger C., James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *Academy of Management Review* 20(3): 709–34.
- McKnight, D. Harrison, Vivek Choudhury, and Charles Kacmar. 2002. Developing and Validating Trust Measures for E-Commerce: An Integrative Typology. *Information Systems Research* 13(3): 334–59.
- Morton, Rebecca B., and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge, UK: Cambridge University Press.
- Munoz, Jordi, Mariano Torcal, and Eduard Bonet. 2011. Institutional Trust and Multilevel Government in the European Union: Congruence or Compensation? *European Union Politics* 12(4): 551–74.

- Nosek, Bryan A., and Daniel Lakens. 2014. Registered Reports: A Method to Increase the Credibility of Published Results. *Social Psychology* 45(3): 137–41.
- Open Science Collaboration. 2015. Estimating the Reproducibility of Psychological Science. *Science* 349(6251): aac4716. <https://doi.org/10.1126/science.aac4716>.
- Pedersen, Morgen J., and Justin M. Stritch. 2018. RNICE Model: Evaluating the Contribution of Replication Studies in Public Administration and Management Research. *Public Administration Review* 78(4): 606–12.
- Peters, B. Guy. 2000. Policy Instruments and Public Management: Bridging the Gaps. *Journal of Public Administration Research and Theory* 10(1): 35–47.
- Pollitt, C. 2008. *Time, Policy, Management*. Oxford: Oxford University Press.
- Rozin, Paul, Amy Wrzesniewski, and Deidre Byrnes. 1998. The Elusiveness of Evaluative Conditioning. *Learning and Motivation* 29: 397–415.
- Schmidt, S. 2009. Shall we Really Do it Again? The Powerful Concept of Replication Is Neglected in Social Sciences. *Review of General Psychology* 13: 90–100.
- Smith, Gareth, and Alan French. 2009. The Political Brand: A Consumer Perspective. *Marketing Theory* 9(2): 209–26.
- Stuart, Elnora W., Terence A. Shimp, and Randall W. Engle. 1987. Classical Conditioning of Consumer Attitudes: Four Experiments in an Advertising Context. *Journal of Consumer Research* 14(3): 334–49.
- Teodoro, Manuel P., and Sueng-Ho An. 2017. Citizen-Based Brand Equity: A Model and Experimental Validation. *Journal of Public Administration Research and Theory* 28(3): 321–38.
- Verlegh, Peeter W.J., and Jan-Benedict E.M. Steenkamp. 1999. A Review and Meta-Analysis of Country-of-Origin Research. *Journal of Economic Psychology* 20(5): 521–46.
- Walker, Richard M., Oliver James, and Gene A. Brewer. 2017. Replication, Experiments and Knowledge in Public Management Research. *Public Management Review* 9(9): 1221–34. <https://doi.org/10.1080/14719037.2017.1282003>.

## Supporting Information

A supplemental appendix can be found in the online version of this article at [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1540-6210](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1540-6210).