The Importance of Personal Vote Intentions for the Responsiveness of Legislators: A Field Experiment

Appendix

Damien Bol, Thomas Gschwend, Thomas Zittel, Steffen Zittlau[1]

[1] Damien Bol, King's College London, damien.bol@kcl.ac.uk; Thomas Gschwend, University of Mannheim, gschwend@uni-mannheim.de; Thomas Zittel, Goethe University Frankfurt, zittel@soz.uni-frankfurt.de; Steffen Zittlau, University of Mannheim and StatistikR.net, info@statistikR.net.

Appendix A. Ethical considerations

Field experiments in which researchers send emails to legislators on behalf of a citizen have important advantages, and thus are more and more common in political science. They however also raise ethical concerns. As a general rule, researchers evidently should not deceive experimental subjects, and should strive for their informed consent. However, we consider field experiments that involve deception important since they allow to pursue conflicting goals that are important and that preclude informed consent. As far as the goals are concerned, we particularly consider it crucial to extract unhindered and objective insights in the basic processes in democratic systems. Under democratic rule, the public has a right to know how they are governed and what kind of biases or blind spots might affect political decision-making. As far as the means are concerned, we consider field experiments crucial to reach these goals. Revealing our real identity for example in a pre-briefing effort would compromise the advantage of randomization, augment the risk to receiving biased reactions from legislators, and thus would result in poor reflections of the actual communications between the constituents and their representatives.

To minimize ethical ramifications, we took several precautions to reduce the harm caused by the experiment on our subjects. First, responding to an email is a task that occurs frequently over the working day of a legislator, even without our intervention (McClendon 2012). Compared to the overall volume of email, our extra email provided a minor nuisance to legislators. Second, in our email we ask a very general question about future projects if re-elected as to minimize time burdens and to also not extract confidential types of information. Third, we completely anonymized the dataset, so that it is no longer possible to identify the individual legislators that took part in our experiment. Fourth, we do not report any effects related to parties to prevent any strategic use of our results in the party-political game and thus to prevent harm to any of the parties involved. Fourth, to further probe our considerations, we subjected our project to an ethics audit at the University of Mannheim. The responsible audit board came to a positive conclusion and found that we conducted an experiment that is relevant for the scientific community and the public and that we took all precautions necessary to minimize harm in a professional way.

Appendix B. Original text of email treatments

From: [Name]@gmx.de Object: Bürgeranfrage: Ihre Arbeit als Abgeordnete[r] Sehr geehrte(r) [Name des/der MdB], Ich heiße [Name] und komme aus [Stadt im Wahlkreis]. Ich sende Ihnen diese E-mail als Bürger, der um seine Zukunft besorgt ist. Normalerweise wähle ich etwas anderes, in letzter Zeit bin ich aber auf Sie und Ihre Arbeit als Abgeordnete[r] aufmerksam geworden. Ich finde es gut wie kompetent Sie sich für die Belange Ihrer Wähler einsetzen. Ich kann mir gut vorstellen, meine Stimme bei der nächsten Wahl Ihnen zu geben. Damit ich eine informierte Wahlentscheidung treffen kann, ist es mir wichtig zu wissen was Sie in der nächsten Zeit machen wollen. Können Sie mir bitte sagen was das wichtigste politische Vorhaben ist, das Sie voranbringen wollen und an dem Sie sich messen lassen wollen bei der nächsten Wahl? Allerbesten Dank schon einmal im Voraus für Ihre Antwort. Mit freundlichen Grüßen, [Name]	From: [Name]@gmx.de Object: Bürgeranfrage: Arbeit der [PARTEI] im Bundestag Sehr geehrte(r) [Name des/der MdB], Ich heiße [Name] und komme aus [Stadt im Wahlkreis]. Ich sende Ihnen diese E-mail als Bürger, der um seine Zukunft besorgt ist. Normalerweise wähle ich etwas anderes, in letzter Zeit bin ich aber auf die Arbeit der [PARTEI] im Bundestag aufmerksam geworden. Ich finde es gut wie kompetent sich Ihre Partei für die Belange ihrer Wähler einsetzt. Ich kann mir gut vorstellen, meine Stimme bei der nächsten Wahl der [PARTEI] zu geben. Damit ich eine informierte Wahlentscheidung treffen kann, ist es mir wichtig zu wissen was Sie in der nächsten Zeit machen wollen. Können Sie mir bitte sagen was das wichtigste politische Vorhaben ist, das Sie voranbringen wollen und an dem Sie sich messen lassen wollen bei der nächsten Wahl? Allerbesten Dank schon einmal im Voraus für Ihre Antwort. Mit freundlichen Grüßen, [Name]

Appendix C. Randomization checks

The analysis below shows that neither the alias (Table C1), nor the wave (Table C2), have a systematic impact on legislator's response behavior.

Table C1. Randomization checks across aliases

Name	No Response	Response	Total
Alexander Müller	38 (31%)	85 (69%)	123 (100%)
Markus Becker	51 (42%)	71 (58%)	122 (100%)
Michael Weber	41 (35%)	77 (65%)	118 (100%)
Thomas Schmidt	53 (40%)	78 (60%)	131 (100%)
Total	183 (37%)	311 (63%)	494 (100%)
N	494		

Note: Entries are absolute number of responses. Row percentages are in parentheses. No systematic differences across names according to a Pearson $\chi 2$-test for the independence of the rows and columns, with $\chi 2$ (3) = 4.1, p = .25

Table C2. Randomization checks across waves

Wave	No Response	Response	Total
1	48 (38%)	77 (62%)	125 (100%)
2	40 (34%)	77 (66%)	117 (100%)
3	52 (42%)	71 (58%)	123 (100%)
4	43 (33%)	86 (66%)	129 (100%)
Total	183 (37%)	311 (63%)	494 (100%)
N	494		

Note: Entries are absolute number of responses. Row percentages are in parentheses. Pearson $\chi 2$-test for the independence of the rows and columns, with $\chi 2$ (3) = 2.71, p = .44

Appendix D. Balance tests

We demonstrate that the covariates in the dataset are orthogonal to our key experimental treatment. Bivariate tests demonstrate that the type of email (personal or partisan vote intention) is unrelated to other characteristics of the email, namely the wave at which it was sent (Table D1) and the sender's name (Table D2). Also, it is not related to the key independent variable, namely the legislator's mode of election (Table D3), as well as two pre-treatment variables, namely gender (Table D4) and age (Table D5).

These results confirm that our covariates do not systematically predict whether someone got assigned a personal or a party representation treatment.

Table D1. Balance test: Waves

Wave	Partisan Vote Intention	Personal Vote Intention	Total
1	60	65	125
	(48%)	(52%)	(100%)
2	57	60	117
	(49%)	(51%)	(100%)
3	61	62	123
	(50%)	(50%)	(100%)
4	64	65	129
	(50%)	(50%)	(100%)
Total	242	252	494
	(49%)	(51%)	(100%)
N	494		

Note: Entries are absolute number of responses. Row percentages are in parentheses. No systematic differences across waves according to a Pearson $\chi 2$-test for the independence of the rows and columns, with $\chi 2$ (3) = .09, p = .993

Table D2. Balance test: Aliases

Name	Partisan Vote Intention	Personal Vote Intention	Total
Alexander Müller	60	63	123
	(49%)	(51)	(100%)
Markus Becker	59	63	122
	(48%)	(52)	(100%)
Michael Weber	58	60	118
	(49%)	(51)	(100%)
Thomas Schmidt	65	66	131
	(50%)	(50)	(100%)
Total	242	252	494
	(49%)	(51%)	(100%)
N	494		

Note: Entries are absolute number of responses. Row percentages are in parentheses. No systematic differences across names according to a Pearson $\chi 2$-test for the independence of the rows and columns, with $\chi 2$ (3) = .04, p = .998

Table D3. Balance test: Mode of Election

Mode of Election	Partisan Vote Intention	Personal Vote Intention	Total
List	118	136	254
	(46%)	(54%)	(100%)
Nominal	124	116	240
	(52%)	(48%)	(100%)
Total	242	252	494
	(49%)	(51%)	(100%)
N	494		

Note: Entries are absolute number of responses. Row percentages are in parentheses. No systematic differences across names according to a Pearson $\chi 2$-test for the independence of the rows and columns, with $\chi 2$ (1) = 1.34, p = .247

Table D4. Balance test: Gender

	Partisan Vote Intention	Personal Vote Intention	Total
Male	166	164	330
	(50%)	(50)	(100%)
Female	76	88	164
	(46%)	(54)	(100%)
Total	242	252	494
	(49%)	(51%)	(100%)

Note: Entries are absolute number of responses. Row percentages are in parentheses. No systematic differences across gender according to a Pearson $\chi 2$-test for the independence of the rows and columns, with $\chi 2(1)$ = .69 and p = .407.

Table D5. Balance test: Age

	Mean (Partisan Vote Intention)	Mean (Personal Vote Intention)	Diff.	Stand. Err.	H: Diff. ≠ 0
Age	51.75	51.58	0.17	(0.85)	0.842

N = 494. No systematic age differences across treatment groups.

Appendix E. Linear probability models

We reproduce the analysis presented in Table 1 of the main text using a linear probability models instead of logit regression models. The results are similar to those of Table 1 (see Table E1).

Table E1. Linear probability models

	(1)	(2)	(3)
Personal Vote Intention	0.08*	0.07*	-0.02
	(0.04)	(0.04)	(0.06)
Mode of Election (1 = Nominal)		0.08	-0.02
		(0.06)	(0.08)
Personal Vote Intention * Mode of Election			0.19**
			(0.09)
Gender (1 = Male)		0.02	0.02
		(0.05)	(0.05)
Age		-0.00	-0.00
		(0.00)	(0.00)
Party Fixed Effects	No	Yes	Yes
Constant	0.59**	0.58**	0.64**
	(0.03)	(0.16)	(0.17)
N	494	494	494

Note: Entries are coefficients estimated from OLS regression models. Standard errors in parentheses. * $p < .10$, ** $p < .05$.

Appendix F. Treatment effects in split samples

We reproduce the analysis presented in the main text in using a split-sample strategy: we separate the sample into two groups, i.e., nominally-elected legislators and those elected in the party-list, and re-estimate the models.

First, we look at the difference in response rates and perform a difference in means t-test. Among nominally-elected legislators (N=240), those who receive a personal vote intention email are more likely to respond than those who receive a partisan vote intention email by 19% points, i.e. 72% of response rate vs. 53% ($p < .01$, t-test difference in means). Among legislators elected via a party list, (N=254) those who receive a personal vote intention email are as likely to respond than those who receive a partisan vote intention email, i.e. 63% of response rate vs. 64% ($p = 0.75$, t-test difference in means).

Second, we reproduce Model 1 of Tables 1, 2, and E1 in using a similar split-sample strategy. The tables below show that the personal vote intention email increases the probability of responding (logit, OLS), and of responding fast (Cox). See Table F1.

Table F1. Split-sample regression models.

	Nominal Legislators (Logit)	Party-list Legislators (Logit)	Nominal Legislators (OLS)	Party-list Legislators (OLS)	Nominal Legislators (Cox)	Party-list Legislators (Cox)
Personal Vote Intention	0.84**	-0.08	0.19**	-0.01	0.50**	-0.01
	(0.28)	(0.26)	(0.06)	(0.06)	(0.17)	(0.16)
Log-Likelihood	-154.0	-166.8			-759.4	-827.6
N	240	254	240	254	240	254

Note: Entries are coefficients estimated from logit regression models (logit), OLS regression models (OLS), and Cox proportional hazard regression models (Cox). Standard errors in parentheses. * $p < .10$, ** $p < .05$.

Appendix G. Power calculations and retrospective design analysis

There is always a possibility that the magnitude and the sign of treatment effects that we estimate are due to chance. We follow the recommended approach of Gelman and Carlin (2014) and perform 'postdata design calculations' (p.643) to evaluate the reliability of the estimated treatment effects given the sample size.

Where is the problem? In our analysis, we have one estimated treatment effect (t^{est}) of the unknown true treatment effect (t^{true}). Suppose that we conduct a hypothetical replication study and obtain a replicated treatment effect (t^{rep}) using a design and sample size identical to the one the original study and assume that the estimated standard error of t^{rep} is the same as the one of t^{est}. In such a context, Gelman and Carlin (2014, p.643) propose three quantities of interest. First, an important quantity is the probability that t^{rep} is larger (in absolute terms) than the critical value defining "statistical significance" in the original study (*Power*). Second, another important quantity is the probability that t^{rep} has the incorrect sign (*Type S error*). Third, a final quantity of interest is the exaggeration ratio, i.e., a ratio calculating by how much the absolute value of t^{rep} overestimates the value of t^{true} in absolute terms (*Type M error*).

Given that this approach requires unbiased and normally distributed treatment effect estimates, we rely on the results of the linear probability models (Appendix E). Moreover, since we show, in the original study, that the treatment effect is entirely driven by the subsample of nominal legislators (see e.g., Figure 2), we focus on this subsample. The t^{est} derived from the linear probability model applied to nominal legislators is the one that we use in the retrospective design analysis, i.e., 0.19 with a standard error 0.06 (see Appendix F).

By definition, t^{true} is unknown. Gelman and Carlin (2014) suggest turning to external information in the literature (as in traditional power analysis) to determine a range of potential values of t^{true}. Our starting point to find comparable studies is the meta-analysis of field experiments aiming at studying the responsiveness of legislators presented in Costa (2017). One key result of the study is that the levels of responsiveness considerably vary across studies. We thus select one that is reasonably close to ours: Broockman (2013) who studies the responsiveness of elected legislators. In this study, he uses two treatments: whether the sender lives in the legislator's district or not, and whether they have a 'putative' black or white name. We compare our treatment effect to Broockman's treatment effect of living in or out of the district. In our study, the email of both the treatment and control group feature a sender living in the legislator's district. Yet, only in the treatment group (*personal vote intention*), there is a clear signal that the voter is intending to vote for the legislator just like a sender living in the district. So, we can see Broockman's treatment as similar to ours.

Broockman's (2013) effect of the treatment living in the district vs. out of the district is +26.6%-points. This is our first reference point. Yet, this treatment effect is probably an over-estimation of ours as, even in our control group (*partisan vote intention*), the sender signals a partisan attachment. Hence, as a second reference point, we reduce Broockman's (2013) sample to the black legislators who received an email from a black sender. In the US context, we can see race as similar to a partisan prime because there is a strong correlation between race and partisan attachment. Looking at Brookman's (2013) subsample of black legislators receiving an email from a black sender, we observe an effect of the treatment living in the district vs. out of the district of +15.6%-points. This is our second reference points.

Table G1 reports the quantities of interest that Gelman and Carlin (2014) for the two reference points presented above, together with the one we estimate in our study (0.190, see above). First, we observe that the probability of these effects having the wrong sign (Type-S error) is essentially zero. Moreover, we also see that given those hypothetical effect sizes our study is in fact well powered. Finally, given the sizes of the respectively estimated exaggeration ratio (Type M error) it is very unlikely likely that our t^{est} is a large overestimation of t^{true}. The factor by which the magnitude might have been exaggerated is not larger than 13.5%.

Table G1. Hypothetical Treatment Effect Sizes in Retrospective Design Analysis.

Hypothetical Treatment Effect	Power	Type S-error	Type M-error
0.266	0.99	0.000	1.004
0.190	0.89	0.000	1.069
0.156	0.79	0.000	1.135

Note. *Power* is the probability that the statistical test correctly rejects the null hypothesis; *Type-S error* is the probability of the sign being in the opposite direction of the one in the original study; and, *Type-M error* is the factor by which the magnitude of the effect size in the original study might be exaggerated.

References

Broockman, David E. (2013). Black Politicians Are More Intrinsically Motivated to Advance Blacks' Interests: A Field Experiment Manipulating Political Incentives. *American Journal of Political Science*, 57(3), 521-536.

Costa, Mia (2017). How Responsive are Political Elites? A Meta-Analysis of Experiments on Public Officials. *Journal of Experimental Political Science*, 4(3), 241-254.

Gelman, Andrew, and John Carlin (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, 9(6), 641-651.

McClendon, Gwyneth H. (2012). Ethics of Using Public Officials as Field Experiment Subjects. *Newsletter of the APSA Experiments Section*, 3(1), 13-20.