

Online Supporting Information for

How Responsive are Political Elites?

A Meta-Analysis of Experiments on Public Officials

Online Supporting Information

SI.1 Defining the Inclusion Criteria

At the most basic level, only randomized controlled trials are included.¹ Second, political elites were the subjects of the experiments. To err on the side of inclusivity, I define “political elite” broadly to cover all government officials at any level, whether elected, appointed, or civil servants. Elites are those who, by their position, are expected to be responsive to citizens. Practically, this ranges from Congress members to bureaucrats at the municipal level. This broad approach reflects the state of research on elite responsiveness, which focuses on officials of different types and at different levels of government, sometimes even within the same study. Taking this broad approach also allows me to test any variability in the findings due to the type of elite studied in any given experiment (I elaborate on this in the main text and in Section SI.3 in this SI.).

Third, the experimental treatment took the form of letters or emails sent by constituents—either real or fictitious—to an official.² The letter or emails requested a response of some type and were manipulated to test various predictors of elite responsiveness. For example, some studies randomized the content of the message to signal party or policy congruence (Grose, Malhotra and Van Houweling, 2015) or whether legislators prioritize service-oriented requests over policy-informed messages (Butler, Karpowitz and Pope, 2012). Others randomized the characteristics of the constituent sender to test whether responsiveness is conditioned by racial or class biases (Carnes and Holbein, 2015; Butler, 2014; Janusz and Lajevardi, 2016,

¹This involves random assignment of the treatment by the researcher, commonly administered in the field, although one survey experiment fit the other inclusion criteria as well (Meng, Pan and Yang, 2014). Excluding the one survey experiment from the analysis does not alter the findings presented in this paper, so I include it as to preserve the full set of relevant studies.

²For the survey experiment, officials were asked about hypothetical emails or letters from constituents.

e.g.)

Finally, I am interested in studies that examine “responsiveness” as the dependent variable. This concept is measured by the rate at which elites send responses to the emails or letters. There are a few studies in which other dependent variables could be considered “responsiveness” –such as responding to fact-checking accountability notices by making fewer inaccurate statements (Nyhan and Reifler, 2014), scheduling a meeting with constituents (Carnes and Holbein, 2015; Chin, 2005; Kalla and Broockman, 2015), or voting in a way that is consistent with the constituent communication (Bergan, 2009; Butler, Nickerson et al., 2011; Chen, Pan and Xu, 2015; Meng and Pan, 2015), but these measures are not comparable to the more common operationalizations of “responsiveness.” In the meta-analysis, I focus on the response outcome because it is the most comparable across studies and the most frequently examined. This is coded as a dummy variable (i.e., received a response or did not receive a response). Some discrepancies exist, however, between studies that distinguish between a response and a “good” response. I therefore also consider “meaningful” responsiveness when available and rely on the operationalization employed in each experiment.

SI.2 Locating Studies

There is not much clarity in the literature on the means by which to gather the population of studies for meta-analyses. Previous meta-analyses in political science used a snowball sampling approach to find studies; that is, they mainly focused on searching top academic journals for published papers, their reference lists, and asking personal contacts for unpublished papers (Lau et al., 1999; Lau, Sigelman and Rovner, 2007). To locate all relevant studies for this analysis, I used a wider variety of search tactics in order to be as comprehensive as possible, including searching library/journal databases, conference proceedings, a pre-registration database, and sending out calls to personal contacts and email listserves.

The initial search for pertinent studies using keywords in Google Scholar yielded over

100 articles, but most were excluded after carefully determining that they did not fit all four conditions for inclusion. I also sent out a call for any working or unpublished experiments to the Experimental Section of the American Political Science Association and the POLMETH listserv, the mailing list of the Society for Political Methodology (which was distributed to 3036 recipients), and the Political Methodology section of the American Political Science Association. These requests yielded multiple new studies over the course of several weeks. Additionally, I searched available conference programs for the annual meetings of the American Political Science Association, Midwest Political Science Association, and Southern Political Science Association. I also searched the Social Science Research Network (SSRN) which contains over 300,000 papers across all social scientific disciplines. Finally, I searched design registrations in the Evidence in Governance and Politics (EGAP) database. The EGAP database contained a few potentially relevant pre-registered experiments, but extensive searches for each suggested that no study had yet resulted from the designs. After thoroughly completing each of these steps, I believe I located every pertinent experiment within a very small margin of error.

The following keywords were used in all database searches:

audit, legislator
audit, politic-
audit, constituent
audit, official
audit, elite
audit, congress
audit, respons-
responsiveness
responsiveness, legislator
responsiveness, elite
responsiveness, experiment
responsiveness, official
responsiveness, congress
responsiveness, politic-
responsiveness, communication

responsiveness, constituent
response, legislator
response, elite
response, experiment
response, official
response, congress
response, politic-
response, constituent
experiment, legislator
experiment, elite
experiment, official
experiment, congress
experiment, communication
experiment, constituent
communication, legislator
communication, elite
communication, experiment
communication, congress
communication, constituent

SI.3 Moderator Variables

In this section, I explain in more detail the moderator variables coded for each study in the meta-analysis.

Response cutoff. Whether or not a public official replied to constituent communication might be a function of logistical elements of the experimental design, like how long researchers waited for a response. If it was reported, I coded the duration of time during which responses were collected in days. If this information was not readily available, I inquired with the authors. I was able to record the time allowed for officials to respond for 31 experiments (out of 41) across 10 academic works. The number of days range from 14 to 280 days with a median of 78. I take the natural log of this variable because I expect decreasing marginal returns for each additional day. In other words, waiting one additional day for a response is likely to have a much larger impact on response rate during the first month of the study than doing so after several months.

Racial/ethnic minority constituent. The purpose of many audit experiments is to test whether public officials are racially biased in whether or how they respond to constituent communication (Butler and Broockman, 2011; Einstein and Glick, 2017; Mendez and Grose, 2014; White, Nathan and Faller, 2015, e.g.). For these studies, I coded the race of the sender in order to test the hypothesis that the estimated effects of communication on elite responsiveness are smaller when the constituent is believed to be a minority. In the U.S., researchers often use putatively white and black (and less frequently, Latino) aliases to examine racial bias, but in other countries the differences were expected to be between White and Turkish names like in Germany (Grohs, Adam and Knill, 2015) or Muslim and non-Muslim names like in China (Distelhorst and Hou, 2014). For this reason, I include a binary indicator for whether the sender is a racial or ethnic “minority” or “non-minority.”³ While treating all minority groups as one homogeneous group is problematic, there are not enough studies that focus on the same race to say anything meaningful about general trends, or to simply measure the combined effect while preserving enough degrees of freedom. Recall that in these studies each condition (i.e. white sender vs. black sender) is treated as a separate experiment so that the response rates can be compared across groups. Of the 28 experiments that explicitly divided senders by race, 14 sent communication from non-minorities and 14 sent from minorities.

Service versus Policy. Another potential moderator is the content of the message. Butler, Karpowitz and Pope (2012) find that state and federal legislative offices are more responsive to constituent requests about service rather than those that focus on policy issues. It is therefore possible that whether constituents requested assistance with constituent service or inquired about policy positions could affect the likelihood of public officials to respond

³While this was obvious in studies meant to test the effect of race, other studies also intentionally used White or non-minority aliases in *all* of their letters/emails in order to avoid biasing their findings. Where this was made explicit, I coded these experiments as having a “non-minority” sender.

in a given experiment. I code each experiment as being either service or policy oriented depending on the focus of the communication. Examples of service-oriented requests are

“Can you tell me how to get unemployment benefits?” (Broockman, 2013) and
“Can you direct me to information about applying for public housing here?”
(Einstein and Glick, 2017).

Examples of policy-oriented requests are

“Can you tell me what is the most important political project that you aim to pursue and according to which you want to be measured in the next election?” (Bol et al., 2015), as well as more elaborate opinions on issues such as “Our laws are being violated left and right by illegal immigrants streaming across the border from Mexico. They disregard our laws to get here. This disrespect for the law is exactly why we should not then turn around and give them citizenship. You must support a strong stance against illegal immigration into this country and not support bills that reward people that break the law...” (Grose, Malhotra and Van Houweling, 2015).

Most studies used service-oriented messages in order to avoid the bias against service-oriented messages found in Butler, Karpowitz and Pope (2012); 34 experiments used service requests and only 7 focused on policy.⁴

Level of government. I also code each experiment for the level of government of the elite subjects as “national” or “sub-national” since it is possible different types of elites have different incentives to respond to constituents. On the one hand, officials at the state and

⁴This is despite the fact that most individuals report contacting their legislator regarding policy issues over requests for help. In fact, less than 6% of respondents answering the 2008 Cooperative Congressional Election Study contacted their representative to get help navigating the federal bureaucracy, while over 91% contacted their representative to express their opinion on an issue (Hickey, 2013).

local level are closer to the smaller constituencies that they represent, and thereby might be more responsive to requests. On the other hand, they often have fewer resources and might therefore be less able to respond at the rate of a national-level office. There are 10 national-level experiments and 31 sub-national. To maintain comparability across different forms of government, I do not differentiate between lower level units of government. To be sure, this is a relatively blunt measure that cannot distinguish between types of legislatures. For example, state legislators in California and New York are far more similar to members of Congress than they are to city council members. There is nevertheless reason to believe that distinctions between sub-national and national political elites can elucidate overall trends in the available data.

Elected. On one hand, we might expect the estimated effects of constituent communication to be larger for elected officials. This follows the intuition and finding in Broockman (2013) that state legislators are more likely to respond where an electoral incentive is present. On the other hand, non-elected officials might deal more frequently with the kind of work that involves responding to constituent communication as part of their civil servant duties. I therefore include a final dummy variable to represent whether the public official is elected or not. Thirty-four experiments are conducted on elected officials whereas 7 experiments are conducted on non-elected officials.

U.S. Indicator. Finally, I include an indicator variable for whether or not the experiment was conducted in the United States. Different norms of representation and constituency service would lead us to expect that response rates would vary in different national contexts under different types of government. While testing the effects of constituent communication on responsiveness in *each* country would be ideal, the next countries with the most frequent experiments are Germany and China with only 3 experiments conducted in each. In order to further investigate the patterns of the other moderator variables, I also present the results from a model of just U.S. experiments later in the Supporting Information.

SI.4 Random-effects models

In the main text, I report on the overall response rate estimated using a random-effects model (Figure 1 in the main text). Random-effects models treat any heterogeneity as purely random because there is not just one identical effect for all studies as in fixed-effect models (Borenstein et al., 2010; Hedges and Vevea, 1998). The random-effects model provides an inference about the distribution of true effects since they can vary from study to study (Viechtbauer et al., 2010). The combined effect in a random-effects model therefore represents the weighted mean of the population of true effects. Findings are weighted by their inverse variance in order to give more weight to studies that measure the dependent variable more precisely. However, these weights are then adjusted based on the overall variance in the size of the effects across studies. The variance in the weights across studies ends up being very small since the between-studies variance (τ^2) is relatively large. I also conduct a robustness check of this weighting technique in the section below that adjusts each finding by the number of subjects in the study divided by the total number of subjects in all studies. In this article, I report using the standard adjustment, instead of the sampling-error adjusted and study-level effect sizes (to adjust for “double counting” experiments in the same study) because the different weighting techniques do not significantly affect the results and the adjustment I use here is most standard in meta-analysis (Viechtbauer et al., 2010).

SI.5 Adjustments and Robustness Checks

Because the N across studies ranges from 108 to 5908, it is important to see if the summary estimate is robust to other adjustments that give more weight to more precise studies. I therefore use another approach to weight studies that has previously been used in meta-analyses (Hunter and Schmidt, 2004; Lau, Sigelman and Rovner, 2007) to see if the results vary depending on the weighting adjustment employed. This adjustment weights each finding

by the number of cases in that study divided by the total number of cases in all studies and does not re-weight the studies by τ^2 (the between-studies variance) like in the random-effects model.

Table SI.1: Summary of Meta-Analysis With Adjustments

| Weighting Technique | | | | | |
|-------------------------------------|-------|------------------------------------|-------|---------------------------|-------|
| Standard τ^2 | | Hunter & Schmidt (2004) | | No double-counting | |
| Effect size | S.E. | Effect size | S.E. | Effect size | S.E. |
| 0.529 | 0.022 | 0.550 | 0.036 | 0.510 | 0.035 |

Note: Standard τ^2 is the estimate reported in the main text

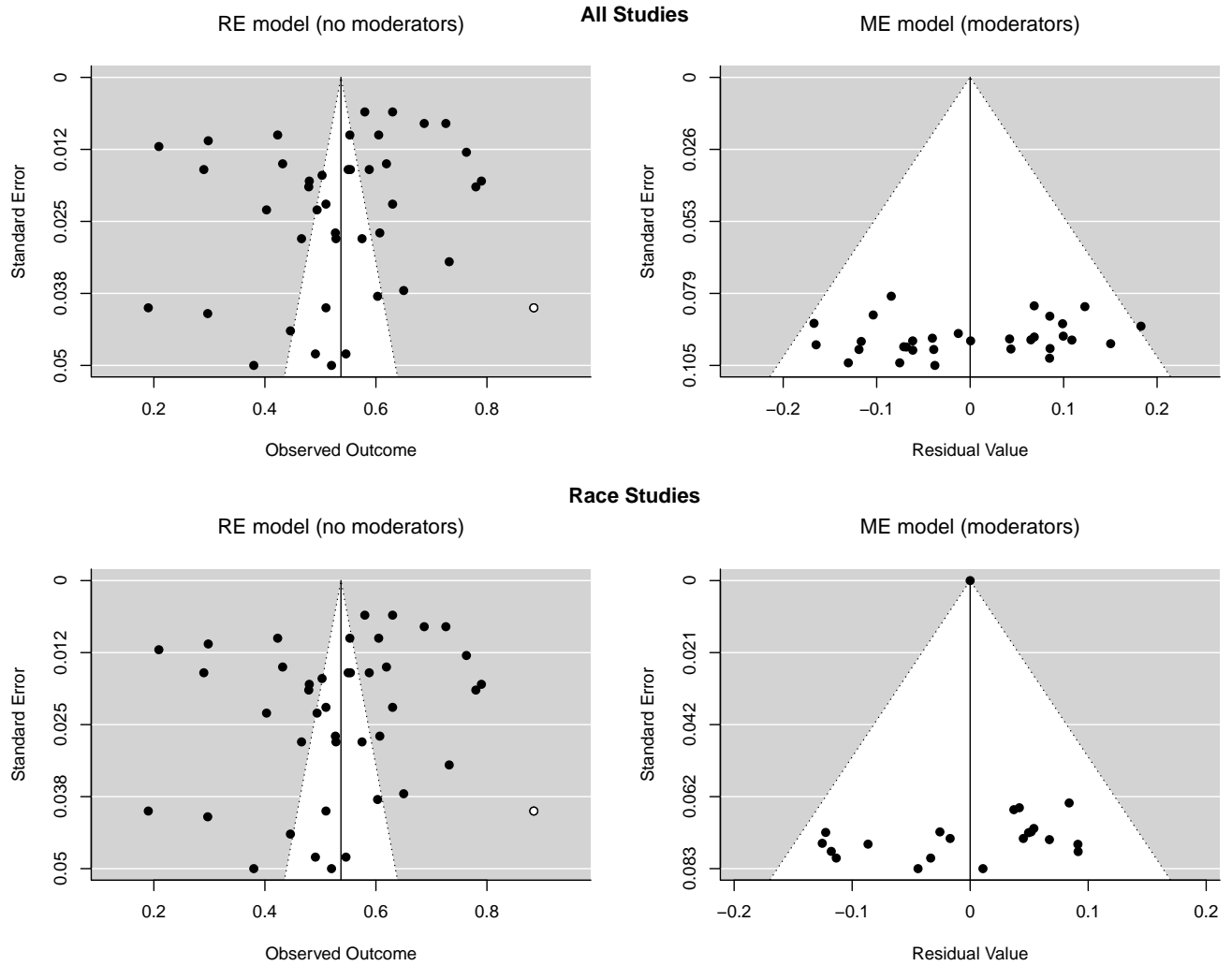
Additionally, I adjust the effect size to combine the findings of experiments that appear in the same paper. In the results reported in the main text, each experiment counts as its own separate case in the meta-analysis. To adjust for “double-counting” these experiments, I calculate the average response rate in each article, weighting for the number of cases on which each experiment is based so that the unit of analysis is the *article or book* instead of experiment.

Table SI.1 shows the summary estimates and standard errors when adjusted by the different weighting techniques described above. The general conclusion is that the different weights do not significantly influence the effect size in either direction.

SI.6 Tests for Publication Bias

Although the summary estimates do not differ for published and unpublished studies, additional tests are often recommended to detect publication bias. In Figure SI.1, the top two figures show funnel plots for the random-effects and mixed-effects models presented in the main text. For the random-effects model, the observed response rates are plotted against their standard errors. The mixed-effects funnel plot shows the difference of each study’s

Figure SI.1: Funnel plots



response rate from the average response rate plotted against their standard errors. This plot could reveal whether studies that find a statistically significant relationship between the moderator variables and responsiveness are more likely to be published. If all the studies were unbiased, there would be a symmetric funnel shape, with studies decreasing in precision increasing in scatter down the y -axis. An asymmetric funnel plot with most of the studies falling on one side of the reference line indicates a relationship between estimates and their

precision (Anzures-Cabrera and Higgins, 2010).

The funnel plots show considerable heterogeneity across findings, in both magnitude and precision, but no evidence of publication bias. Regression tests for funnel plot asymmetry confirms the absence of a relationship between the observed outcomes and their standard errors. ($z = -1.6612$, $p = 0.0967$ for the main mixed effects model) (Viechtbauer et al., 2010).

In order to detect publication bias specifically for the main causal effect the literature has focused on, the effect of race on responsiveness, the bottom two figures plot the same relationships but only for the studies that specifically focus on the race/ethnicity of constituents. This allows us to see whether studies that find a statistically significant difference between minority senders and non-minority senders are more likely to be published. As with the first two plots, no evidence of publication bias is apparent for these studies.

SI.7 Regression Model Without Response Cut-Off Variable

In Table 2 in the main text, I present a mixed-effects model that presents the effect of six moderator variables on elite responsiveness. Since the value for one of those variables, response cut-off, is missing for ten studies in the dataset, I exclude it in a second model here in order to increase power and include the full set of studies in the analysis. Model 1 is reproduced from the main text and Model 2 presents the model without the response cut-off variable. As can be seen here, excluding this variable does not change the statistical or substantive significance of the results presented in Model 1.

SI.8 Regression Models for Studies in the United States

The combined response rate in the United States remains relatively unchanged using the random-effects model without moderators (0.536). Table SI.3 shows the regression estimating the effects of the moderator variables on response rates in the U.S. studies.

Table SI.2: Meta-Regression Analysis Estimating the Effect of Moderators on Elite Responsiveness, with Model 2 added to preserve N

| | (1) | (2) |
|--|------------------|---------------------|
| | Full model | W/o response cutoff |
| Response cutoff <i>(log days)</i> | -.004 (.031) | |
| Minority constituent <i>(baseline=Non-minority)</i> | -.094* (.047) | -.119* (.049) |
| Service <i>(baseline=Policy)</i> | .003 (.071) | .053 (.067) |
| Sub-national <i>(baseline=National)</i> | -.061 (.055) | -.027 (.055) |
| Elected | -.181* (.055) | -.174* (.056) |
| In U.S. | .054 (.054) | .061 (.053) |
| Intercept | .762* (.144) | .670* (.079) |
| Observations | 31 | 41 |
| R^2 | .489 | .263 |
| τ^2 | .011 (.003) | .015 (.004) |

Note: * $p < 0.05$. Model 1 is reproduced from the main text, and excludes 10 studies for which the response cutoff is not reported. Model 2 excludes that variable so all studies can be included. τ^2 represents the amount of heterogeneity among the true effects that is not already accounted for by the moderators. τ^2 estimator: Restricted maximum-likelihood estimation.

Table SI.3: Meta-Regression Analysis Estimating the Effect of Moderators on Elite Responsiveness In U.S. Studies

| | (1) | (2) |
|--|-----------------|---------------------|
| | Full model | W/o response cutoff |
| Response cut-off <i>(log days)</i> | -.008 (.034) | |
| Minority constituent <i>(baseline=Non-minority)</i> | -.087 (.047) | -.107* (.047) |
| Service <i>(baseline=Policy)</i> | .036 (.073) | .117 (.071) |
| Sub-national <i>(baseline=National)</i> | -.038 (.052) | -.034 (.053) |
| Elected | -.097 (.060) | -.115* (.057) |
| Intercept | .647 (.168) | .606* (.084) |
| Observations | 24 | 33 |
| R ² | .326 | .177 |
| τ^2 | .009 (.003) | .011 (.003) |

Note: *p<0.05. Model 1 excludes 9 studies for which the response cutoff is not reported. Model 2 excludes that variables so all studies can be included. τ^2 represents the amount of heterogeneity among the true effects that is not already accounted for by the moderators. τ^2 estimator: Restricted maximum-likelihood estimation.

First, note that while the coefficients again do not change much between models, including more studies allows for more precision in estimating the effects. Therefore, both coefficients for minority sender and elected official are statistically significant in Model 2. Minority constituents in the U.S. are over 10 percentage points less likely to receive a response from public officials than white constituents.

SI.9 Response Quality

Table SI.4: Population of “Good Response” Definitions

| Study | Content | Tone |
|-------------------------|---|------|
| Bishin and Hayes (2016) | meaningful – “The variable is coded dichotomously, where a 1 equals any response that would have led to information that would have proved helpful to the constituent’s request. Therefore, any response that provided information (such as the healthcare.gov website) or specific information about qualifications for coverage are coded as 1. Form responses (simply indicated the legislator had received the email request) or position statements (for or against the health care law) were not coded as meaningful responses– mainly because such responses did not provide service to a constituent with a specific request.” p. 12 | NA |
| Broockman (2013) | helpful – “I coded emails as helpful if they (1) provided the website, email address, physical address, or telephone number of a person or agency that could help a person register for unemployment benefits or (2) invited further contact from the alias in order to provide this information (some replies that ask Tyrone for his phone number so that the legislator could call him).” p. 2 of Supplementary Information | NA |
| Butler (2014) | answered question – “measures whether the official’s response answered the question that was asked; those who did not respond at all and those who responded but did not answer the question are coded the same way.” p. 30 | NA |

| | | |
|------------------------------|---|---|
| Carnes and Holbein (2015) | <p>quality – “the number of characters in the legislators’ reply emails. [If legislators did not respond, this measure took on a value of 0. If legislators sent more than one reply email, we simply summed the character counts for all of them.]”</p> <p>helpful – “provided detailed instructions about where Joey could go to register to vote (in the first experiment) or that offered to schedule a meeting with Joey (in the second experiment.)” p. 15-16</p> | <p>friendly, personal – “an indication that the email was from the legislator herself and not an assistant, a thank you, an offer to provide follow-up help, and/or encouragement to register.” p. 16 (alternate specification, but results not reported in paper)</p> |
| Einstein and Glick (2017) | NA | <p>friendly – “whether the emailer is addressed by proper name. We were lenient in coding ‘yes.’ A named salutation could be as casual as ‘Hi Brett’ or as formal as ‘Dear Ms. Martinez’.” p. 13</p> |
| Grohs, Adam and Knill (2015) | <p>completeness – “responses were coded according to their informational content and the presentation of information. All requests comprised two thematic blocks with different subquestions. If all subquestions of both thematic blocks were answered, a score of 4 points was given.”</p> <p>response quality – “Congruently, the comprehensibility and preparation of the responses were both rated with two points. The maximum score for the subcategory response quality thus amounted to 8 points.”</p> | <p>service orientation – “We gave up to 3 points with regard to the thoroughness of the response, a friendly and courteous tone of the response, and the mentioning of additional contact persons for further questions.” p. 4</p> |
| McClendon (2016) | <p>answered – “Additionally, in order to gauge the effort politicians put into replying, I also coded ‘answered’ as a 1 if the politician supplied the requested information directly or provided the contact information for the bureaucrat, through a carbon copy. ‘Answered’ was coded zero if the politician did not reply or replied only to ask for more information.” p. 7</p> | NA |

White,
Nathan and
Faller
(2015)

absolutely accurate – “replies that provided links to state websites with official instructions about voting requirements.” p. 13 of Supplementary Information

friendly – “Emails marked as ‘friendly’ contained ‘explicitly friendly language, such as use of the senders name in the salutation or sign-off. Examples included ‘Dear (name),’ ‘Let us know if you have any more questions’ and ‘Have a great day.’ p. 36

Note: Studies are listed in ascending alphabetical order. The conceptual definition that the authors (explicitly as stated in the paper) attempt to measure are in **bold**; the according measurement and/or coding scheme are taken verbatim from the paper and are in quotations.

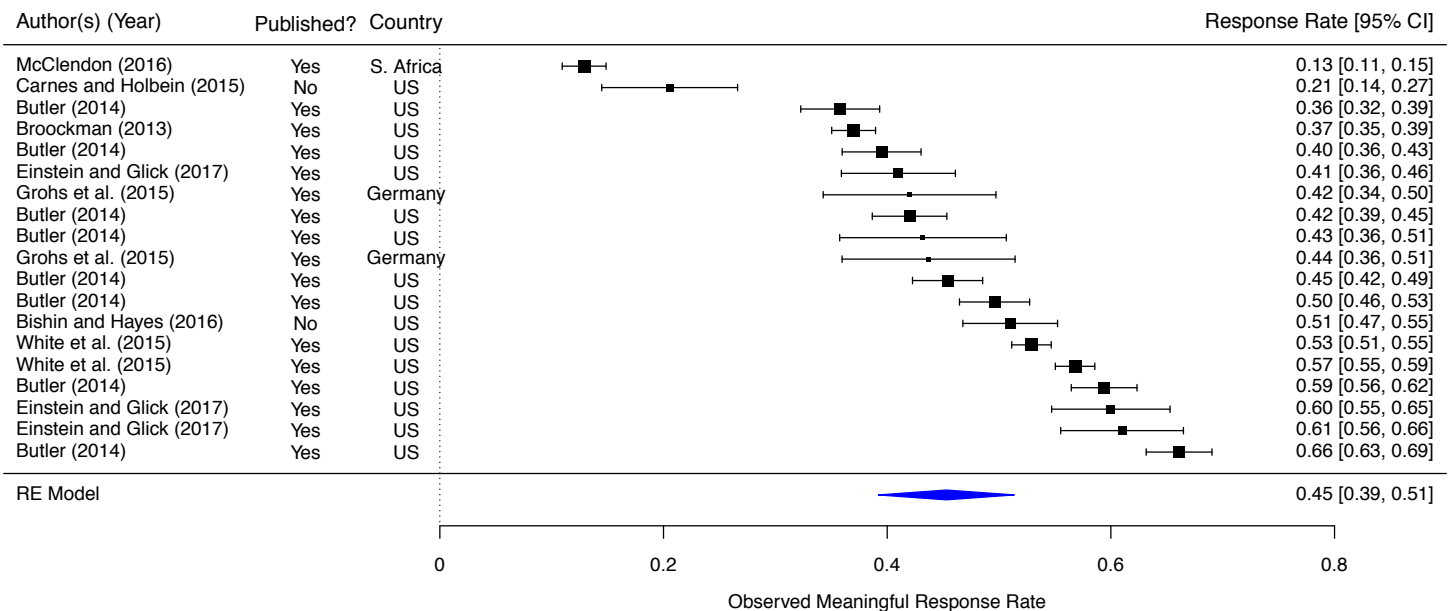
In the main text, I focus on explaining differences that influence the rate at which political elites will respond to constituent communication. I focus less on the average rate of response. But it is important to note that some studies additionally measure the *quality* of the response received, so I therefore consider that alternative outcome variable here. How often do political elites respond *well* to constituent communication? To investigate this question, I searched for cases where studies not only measured whether or not their letters were responded to, but how many quality responses they received as well.

There are many different ways the literature has operationalized a “good” response. For example, some definitions focused on the accuracy of the response (White, Nathan and Faller, 2015) where others focused on attempts to be helpful (Carnes and Holbein, 2015) or friendly (Einstein and Glick, 2017). See the table in this document for the population of definitions and coding schemes used for this variable. Since definitions vary from study to study, I rely on whatever operationalization was employed in a given study rather than impose a definition of my own.⁵ A common proposition underlies each operationalization: that whether or not a public official responds at all only matters insofar as that response is meaningful to the constituent.

⁵While it’s possible these differences in operationalization can be an additional source of heterogeneity, among the 8 studies for which both measures are available, there is actually a substantial decrease in the test for heterogeneity among the effects using the meaningful response variable compared to the binary response variable ($Q = 3695.39$ and 1816.6578 , $df = 40$ and 18 , $p < .0001$, respectively).

Given that only 19 studies out of the 41 measured the quality of responses, I only present the combined effect estimated using a random effects model without moderators. A forest plot of the observed effect sizes is presented in Figure SI.2 with the summary estimate at the bottom. The main effect of constituent communication on receiving a *meaningful* response from a public official is 0.453, a difference of 8 points from receiving any response (0.53), meaningful or not.⁶

Figure SI.2: Forest Plot of Meaningful Responsiveness



Note: Figure plots the estimate and 95% confidence interval for each study. Estimates are represented by the black boxes and sized proportional to their precision. Studies with larger boxes are given more weight in the calculation of the effect size.

This finding is somewhat surprising given the literature on responsiveness and represen-

⁶However, to ensure that this difference is not due to some sort of bias in the subset of studies for which a meaningful response rate is available, I also compared the meaningful response rate to the base response rate among that subset of studies. The base response rate in that set of studies is 0.582, which based on a *z*-test for difference in coefficients, does not significantly differ from the base response rate in all studies. Thus this difference is not a function of comparing different samples.

tation in general. First, we would expect the overall proportion of meaningful responses to be much smaller. After all, only insofar as communication is felt to be meaningful by constituents does it link symbolic and substantive representation, thereby increasing trust and satisfaction with the representational relation. This is usually treated as a fairly high bar to meet (See, e.g., Eulau and Karps 1977). Additionally, responses from elected officials are commonly algorithmically and automatically generated (Fitch, Goldschmidt and Cooper, 2011). It would presumably be an error to consider these generic, form messages meaningful or satisfying to the constituent recipient. We might therefore expect there to be a bigger difference between the basic response rate and meaningful response rate. Yet the 8 point difference is not statistically significant and could be considered relatively small, which suggests that most responses *are* in fact “meaningful.” If constituents are going to hear back from a public official, there is a good chance they will be satisfied (as defined by researchers) with the quality of the response.

Then again, perhaps the lack of difference is due to the weak conceptualization of this variable. For instance, some scholars considered a response good depending on the *tone* of the response (friendly, personal, etc.), while others prioritized the *content* (if it answered the constituent’s question, etc.). These two takes on what constitutes a good response rely on different assumptions about responsiveness. While the latter more closely follows democratic theory in its presupposition that elites are responsive if they represent constituents’ needs (Urbinati and Warren, 2008, e.g.), the former’s focus on tone and friendliness implies that being responsive is not necessarily the same as following through on a request; rather, constituents most like “being heard.” It is possible that this more lenient definition of response quality biases the meaningful response rate upwards and closes the gap between how many political elites responded to constituent communication and how many political elites were considered to have responded *well*.

References

- Anzures-Cabrera, Judith and Julian PT Higgins. 2010. “Graphical displays for meta-analysis: An overview with suggestions for practice.” *Research Synthesis Methods* 1(1):66–80.
- Bergan, Daniel E. 2009. “Does grassroots lobbying work? A field experiment measuring the effects of an e-mail lobbying campaign on legislative behavior.” *American Politics Research* 37(2):327–352.
- Bishin, Benjamin and Thomas Hayes. 2016. “Do Elected Officials Service the Poor? A Field Experiment on the U.S. Congress.” Paper presented at the annual Southern Political Science Associate conference, San Juan, Puerto Rico (January 9, 2016). <https://www.dropbox.com/s/ahhctobf0gb1vbc/Do%20Elected%20officials%20Service%20the%20Poor%3F%20%20A%20Field%20Experiment%20on%20Congress..pdf?dl=0>.
- Bol, Damien, Thomas Gschwend, Thomas Zittel and Steffen Zittlau. 2015. “The Electoral Sources of Good Political Representation A Field Experiment on German MPs.” Paper presented at the Annual Meeting of the European Political Science Association, Vienna (June 25–27, 2015). <https://www.dropbox.com/s/vw7682r0qwto85b/The%20Electoral%20Sources%20of%20Good%20Political%20Representation-A%20Field%20Experiment%20on%20German%20MPs.pdf?dl=0>.
- Borenstein, Michael, Larry V Hedges, Julian Higgins and Hannah R Rothstein. 2010. “A basic introduction to fixed-effect and random-effects models for meta-analysis.” *Research Synthesis Methods* 1(2):97–111.
- 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.

- Butler, Daniel M. 2014. *Representing the Advantaged: How Politicians Reinforce Inequality*. Cambridge University Press.
- Butler, Daniel M, Christopher F Karpowitz and Jeremy C Pope. 2012. “A field experiment on legislators’ home styles: service versus policy.” *The Journal of Politics* 74(02):474–486.
- Butler, Daniel M and David E Broockman. 2011. “Do politicians racially discriminate against constituents? A field experiment on state legislators.” *American Journal of Political Science* 55(3):463–477.
- Butler, Daniel M, David W Nickerson et al. 2011. “Can Learning Constituency Opinion Affect How Legislators Vote? Results From a Field Experiment.” *Quarterly Journal of Political Science* 6(1):55–83.
- Carnes, Nicholas and John Holbein. 2015. “Unequal Responsiveness in Constituent Services? Evidence from Casework Request Experiments in North Carolina.” Working Paper. http://people.duke.edu/~nwc8/carnes_and_holbein.pdf.
- Chen, Jidong, Jennifer Pan and Yiqing Xu. 2015. “Sources of Authoritarian Responsiveness: A Field Experiment in China.” *American Journal of Political Science* 60(2):383–400.
- Chin, Michelle L. 2005. “Constituents Versus Fat Cats Testing Assumptions About Congressional Access Decisions.” *American Politics Research* 33(6):751–786.
- Distelhorst, Greg and Yue Hou. 2014. “Ingroup bias in official behavior: A national field experiment in China.” *Quarterly Journal of Political Science* 9(2):203–230.
- Einstein, Katherine Levine and David M Glick. 2017. “Does Race Affect Access to Government Services?: An Experiment Exploring Street Level Bureaucrats and Access to Public Housing.” *American Journal of Political Science* 61(1):100–116.

- Eulau, Heinz and Paul D Karpis. 1977. "The puzzle of representation: Specifying components of responsiveness." *Legislative Studies Quarterly* pp. 233–254.
- Fitch, Brad, Kathy Goldschmidt and Nicole Folk Cooper. 2011. "Communicating with Congress: How Citizen Advocacy is Changing Mail Operations on Capitol Hill." *Congressional Management Foundation* .
- Grohs, Stephan, Christian Adam and Christoph Knill. 2015. "Are Some Citizens More Equal than Others? Evidence from a Field Experiment." *Public Administration Review* .
- Grose, Christian R., Neil Malhotra and Robert P. Van Houweling. 2015. "Explaining Explanations: How Legislators Explain Their Policy Positions and How Citizens React." *American Journal of Political Science* .
- Hedges, Larry V and Jack L Vevea. 1998. "Fixed-and random-effects models in meta-analysis." *Psychological Methods* 3(4):486.
- Hickey, Emily Grace. 2013. "Essays in Congressional Communication." Doctoral Dissertation, Harvard University. https://dash.harvard.edu/bitstream/handle/1/11156822/Hickey_gsas.harvard_0084L_10787.pdf?sequence=1.
- Hunter, John E and Frank L Schmidt. 2004. *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.
- Janusz, Andrew and Nazita Lajevardi. 2016. "Differential Responsiveness: Do Legislators Discriminate Against Hispanics?" Paper presented at the annual Midwest Political Science conference, Chicago, Il (April 5, 2014). updated version: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2799043.
- Kalla, Joshua L and David E Broockman. 2015. "Campaign Contributions Facilitate Ac-

- cess to Congressional Officials: A Randomized Field Experiment.” *American Journal of Political Science* .
- Lau, Richard R, Lee Sigelman, Caroline Heldman and Paul Babbitt. 1999. “The effects of negative political advertisements: A meta-analytic assessment.” *American Political Science Review* 93(04):851–875.
- Lau, Richard R, Lee Sigelman and Ivy Brown Rovner. 2007. “The effects of negative political campaigns: a meta-analytic reassessment.” *Journal of Politics* 69(4):1176–1209.
- McClendon, Gwyneth H. 2016. “Race and Responsiveness: An Experiment with South African Politicians.” *Journal of Experimental Political Science* pp. 60–74.
- Mendez, Matthew and Christian Grose. 2014. “Revealing Discriminatory Intent: Legislator Preferences, Voter Identification, and Responsiveness Bias.” USC CLASS Research Paper No. 14-17. <http://ssrn.com/abstract=2422596>.
- Meng, Tianguang and Jennifer Pan. 2015. “Responsive to Whom? A Survey Experiment of the Influence of Superiors, Businesses, and Residents on China’s Subnational Officials.” Paper presented at the Annual Meeting of the American Political Science Association, San Francisco, CA (September 3-6, 2015). <https://www.dropbox.com/s/vv0vmufuwxfb9wz/Responsive%20to%20Whom%3F%20A%20Survey%20Experiment%20of%20the%20Influence%20of%20Superiors%2C%20Businesses%2C%20and%20Residents%20on%20Chinas%20Subnational%20Officials.pdf?dl=0>.
- Meng, Tianguang, Jennifer Pan and Ping Yang. 2014. “Conditional Receptivity to Citizen Participation Evidence From a Survey Experiment in China.” *Comparative Political Studies* p. 0010414014556212.
- Nyhan, Brendan and Jason Reifler. 2014. “The Effect of Fact-Checking on Elites: A Field Experiment on US State Legislators.” *American Journal of Political Science* .

- Urbinati, Nadia and Mark E Warren. 2008. "The concept of representation in contemporary democratic theory." *Annual Review of Political Science* 11:387–412.
- Viechtbauer, Wolfgang et al. 2010. "Conducting meta-analyses in R with the metafor package." *Journal of Statistical Software* 36(3):1–48.
- White, Ariel R, Noah L Nathan and Julie K Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109(01):129–142.