

Improving Measures of Responsiveness for Elite Audit Experiments *

Mia Costa †

April 21, 2017

Abstract

Scholars define elite “responsiveness” to constituent communication in a multitude of ways. Yet there is not much consistency in the literature on measuring the quality of responsiveness to constituent contact, nor do we know how constituents themselves experience and evaluate this contact. In this paper, I conduct three tests to examine how individuals evaluate communication with elected officials to ultimately increase the external validity of measures of elite responsiveness used in audit experiments. Based on the findings, I develop a new weighted measure for responsiveness and demonstrate its use through an application. I conclude by providing concrete suggestions for researchers conducting future audit studies on elite responsiveness.

*Prepared for New Faces in Political Methodology IX at Pennsylvania State University, April 29, 2017. I would like to thank Dan Butler for providing replication data and for helpful advice. I would also like to thank Thomas Kennedy and Anthony Rentsch for excellent research assistance. All errors are my own.

†micosta@polsci.umass.edu; Department of Political Science, University of Massachusetts Amherst

1 Introduction

Experimental audit studies have become an increasingly popular methodological approach to understand the causal mechanisms behind social phenomena, typically discrimination (Butler and Crabtree, 2016; Vuolo, Uggen and Lageson, 2016). For example, audit studies have been used to measure the extent to which bias exists in employment and hiring practices (Pager, 2007), housing (Bertrand and Mullainathan, 2004; Einstein and Glick, 2016), and health provisions (Burgess et al., 2007). In political science, audit studies have burgeoned in the last decade with the aim of investigating how responsive governments are to constituent service requests (Costa, 2016). These studies involve randomizing the type of communication sent to public officials, evaluating how many public officials respond, and comparing the response rates across groups. Typically, the aim is to determine whether elected officials are more responsive to some constituents than others, and most scholars operationalize “responsiveness” as whether government officials respond at all and code it as a binary variable (i.e., responded or did not respond).

More recently, scholars have been careful to distinguish between any response and a “good” response (See, e.g. Broockman, 2013; Butler, 2014; Einstein and Glick, 2016; McClen- don, 2015; White, Nathan and Faller, 2015). The general sentiment consistent across these studies is that some responses are meaningful while others merely satisfy some minimum required effort on behalf of the public official. If responsiveness is meant to capture something about representation more generally, the quality of the response is crucial. Whether or not a public official responds at all only matters insofar as that response is satisfying to the constituent.

This more nuanced approach to responsiveness takes varied forms of measurement in the literature. Depending on the study in question, a good response is conceptually defined as one that is either “accurate”, “complete”, “helpful”, “friendly”, “welcoming”, “personal”,

“service-orientated”, and so on. I sort some of the ways scholars have measured responsiveness in the literature in Table A1. Variations abound. For example, some studies consider responses friendly or personal if they include a named salutation (“Hi Rosa”, “Dear Tremayne”, “Hello Katherine”, etc.) (Einstein and Glick, 2016); others consider responses friendly only if they include additional friendly language, such as inviting them to follow up with additional questions and thanking the constituents for writing (See Carnes and Holbein, 2015; Grohs, Adam and Knill, 2015; White, Nathan and Faller, 2015).

Some studies additionally measure whether or not the legislator responds in a timely fashion (See Bol et al., 2015; Butler, 2014; Chen, Pan and Xu, 2015; Einstein and Glick, 2016). Timeliness signals an important element of responsiveness; the quicker legislators are to respond, the more attention and effort they afford to those constituents. Yet there is not a clear time after sending a constituent request at which point legislators should be penalized for responding “slowly.” Measures of timeliness in the literature range from 24 hours to 15 days. For example, Einstein and Glick (2016) examine whether responses were received within the first day they were sent out and find that responses from public housing officials to white and black citizens do not statistically differ in their timeliness. Using 2 weeks for this window of time, Butler (2014) finds that non-Latino white legislators are more likely to respond in a timely manner to white constituents than black or Latino constituents.

The lack of congruency among conceptions of what qualifies as a “good” response generates several challenges. First, without a clear theoretical consensus on what a good response should look like, the implications of any given study on elite responsiveness at large are very limited. Second, it is possible that the citizens who are actually involved in this communication exchange with their elected officials have different ideas of what should constitute a good response, challenging the external validity of audit studies. Despite the increasing scholarly focus on constituent communication with political elites, little research has been done in the way of examining the expectations constituents themselves have about this communication.

In this paper, I address this gap by conducting three tests to examine what individuals actually want from communication with elected officials. Overall, I find that the friendliness of legislators’ responses strongly influence individuals’ perceptions of the response’s quality, as well as its timeliness, length, and whether it actually answered the constituent’s question. Other characteristics that previous scholars have used to measure response quality, such as thanking the constituent for writing or asking for more information, did not have a statistically significant effect on individuals’ perceptions of the response. Based on these findings, I develop an alternative outcome measure of responsiveness and apply this measure to a replication of an audit experiment in Butler (2014). I conclude by providing a few concrete suggestions on measurement for researchers conducting audit studies on elite responsiveness.

2 Data and Methods

To examine how people evaluate responses from elected officials, I conducted three separate tests. The first two were survey experiments fielded by YouGov, America using nationally representative internet samples of 1,000 American adults. The first experiment was a part of the University of Massachusetts Amherst module of the 2016 Cooperative Congressional Election Study (CCES) pre-election common content and was fielded between September 28th and November 3rd, 2016. The second experiment was fielded between March 24th and April 1st, 2017.

In both experiments, subjects were first prompted with a vignette in which a constituent emails his or her state legislator to ask how to register to vote. The two experiments differed in one key way. In the first experiment, the constituent in the vignette is named “Jake.” Jake is one of the most common names political scientists have used in audit studies using fictitious constituents. The second experiment uses “Jane” and allows me to test the robustness of the effects and rule out any gendered dynamics.

In both experiments, subjects were first presented with the prompt: “Imagine [Jake/Jane] just moved to a new area. [He/She] emailed [his/her] state legislator asking for information on how to vote. Below is the response [he/she] received from [his/her] state legislator after X days” with X being a randomly generated integer between 1 and 30. Framing the vignette in this way helps to increase the external validity of subjects’ evaluations; state legislators are the most frequent subject type in audit experiments on elite responsiveness, and service-oriented requests, like asking for information on registering to vote, are also highly prioritized over emails about policy (Costa, 2016).

After the prompt, subjects were presented with the (hypothetical) email response from the legislator and asked to evaluate it.¹ Three treatment variables were randomized across the responses shown to respondents.² First, the response either provided an answer to the question or contact information for another office. Some scholars consider providing contact information a helpful response (See, e.g., Broockman 2013) whereas others discount it (See, e.g., McClendon 2015); varying this aspect in the responses allows me to test these assumptions directly. The “answer” email and the “contact information” email are relatively the same length to control for perceived effort the legislator exerted to respond. In addition, the responses varied in whether they were “friendly”: the friendly responses start with a named greeting (specifically, “Dear Jake/Jane”) and end with an invitation to follow-up with additional questions (“Let me know if you have additional questions.”). Finally, the number of days until the state legislator was the randomly generated integer from 1-30 included in the prompt. After viewing the legislator’s response, subjects were then asked to rate the response on its overall quality, friendliness, and helpfulness on a scale from 0-100.

In addition to the two survey experiments, a third test involves raw email data from a 2010 audit experiment conducted on political elites in Butler (2014). The purpose of the

¹See the Appendix for images of the emails shown in each experimental vignette.

²A fourth variable that randomized whether the state legislator was male or female is not included in this analysis.

experiment was to test whether non-Latino white state legislators are biased in their responsiveness to blacks and Latinos. Two research assistants coded 400 of the emails received from state legislators in the experiment for characteristics such as whether it was automated or personalized, included a named salutation, thanked the constituent, included a link to a website, asked for more information, invited the constituent to follow up with further questions, ended with a “sign off” (Sincerely, Best, Regards, etc.), and whether it was sent by the legislator him or herself or a legislative staff member. I manually resolved any disagreements between coders for ultimate coder agreement of 100%. Emails were then anonymized so that the name, email address, and location of elected officials remained confidential. Descriptive statistics of these variables are presented in Table 1. Once anonymized, subjects from Amazon’s Mechanical Turk (MTurk) were recruited to evaluate the emails.

Table 1: Descriptive Statistics of Legislator Emails

	Freq. (#)	Percent (%)
Automated	57	14.3
Rep. (not staff)	202	50.5
Named greeting	243	60.8
Contact info	61	15.2
Website	166	41.5
Thank you	179	44.8
Ask more info	81	20.3
Invite follow up	129	32.3
Sign-off	171	42.8
	Mean (s.d.)	Median
Words	91.7 (141)	58
Characters	480.8 (880.7)	284

The study was conducted on MTurk through April 12th and 13th, 2017.³ Respondents had two disqualifying questions regarding age (must be over 18) and residency (must be residents of the United States) and were paid 50 cents upon completion of the survey. Re-

³For more on MTurk, see Berinsky, Huber and Lenz (2012).

spondents first answered a very brief demographic/political battery before being asked to evaluate 5 email exchanges between constituents and state legislators, chosen randomly from the batch of 400. They were asked to rate each email on a scale from 0-100 based on how satisfied they would be with the response if they received it from their legislator. All respondents were also shown a sixth email that served as an attention check; instructions to give this email a satisfaction rating of 100 were embedded in two separate places in the body of the email. After they evaluated the emails, respondents were presented with an open text box in which they could write in their own words what makes a good email response from an elected official. I recruited a total of 1,000 subjects so that each legislator response would be rated an average of 12.5 times. From those ratings, I produced an average quality measure for each email. I then examine the independent effect of each response characteristic on perceptions of response quality.

3 Results

3.1 Experiment 1 and 2

I report the results of both survey experiments in this section.⁴ Tables 2 and 3 present the effects of the treatments on the dependent variables (evaluations of overall quality, friendliness, and helpfulness) estimated using ordinary least squares. The models include two indicator variables for whether the email response included an answer to the constituent’s question rather than contact information for someone else and included a named greeting and invite to follow up, as well as a variable for the number of days until the state legislator responded to the constituent request (which was a randomly assigned integer from 1-30 days).

Overall, individuals mostly care about the friendliness of the email; the only consistently

⁴The mean evaluations of response quality, friendliness, and helpfulness by condition are shown in Figure A1.

Table 2: Effects of Random Treatment Variables on Email Evaluations: 2016 CCES Experiment

	<i>Dependent variable:</i>		
	Overall Quality	Friendliness	Helpfulness
	(1)	(2)	(3)
Greeting + follow up	9.023* (1.582)	11.189* (1.587)	6.245* (1.675)
Answer vs. contact info	0.058 (1.581)	-0.299 (1.587)	-0.322 (1.674)
Days until response	-0.357* (0.090)	-0.123 (0.091)	-0.190* (0.096)
Constant	53.456* (1.944)	47.251* (1.948)	58.206* (2.055)
Observations	992	994	995
R ²	0.048	0.050	0.018
Adjusted R ²	0.045	0.047	0.015

Note: *p<0.05; Standard errors are in parentheses.

Table 3: Effects of Random Treatment Variables on Email Evaluations: 2017 Experiment

	<i>Dependent variable:</i>		
	Overall Quality	Friendliness	Helpfulness
	(1)	(2)	(3)
Greeting + follow up	6.130* (1.694)	9.971* (1.719)	5.418* (1.772)
Answer vs. contact info	-0.707 (1.693)	-1.800 (1.718)	-2.538 (1.771)
Days until response	-0.296* (0.098)	-0.021 (0.100)	-0.225* (0.103)
Constant	53.001* (2.110)	46.875* (2.141)	60.487* (2.208)
Observations	996	996	996
R ²	0.022	0.034	0.016
Adjusted R ²	0.019	0.031	0.013

Note: *p<0.05; Standard errors are in parentheses.

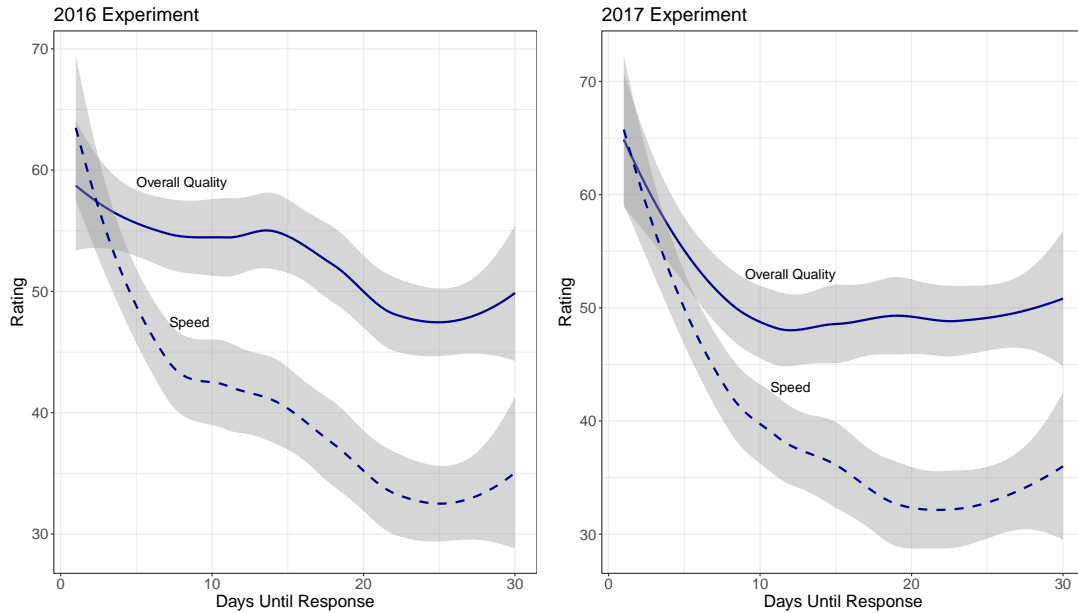
significant predictor of response evaluations is whether the email included a named greeting and invitation to follow-up. This effect is robust across both experiments. If those “friendly” items were included in the legislator’s response, it was rated 9 points higher in overall quality, over 11 points more friendly, and 6.3 points more helpful in the first experiment. In the second experiment, the treatment effects slightly decrease in magnitude but remain statistically significant. Note that for these experiments, it is unclear whether it is the combined effect of the named greeting and the invitation to follow up, or one or the other that is driving the effect. However, I separate these factors in the subsequent MTurk study.

The content of the email –that is, whether it answered the question or pointed the constituent somewhere else– had no statistically significant effect on any perceptions of the legislator’s response in either experiment. This is especially notable since some audit studies do not consider anything but a full answer to the constituent request a meaningful response (See, e.g. Bishin and Hayes, 2016; Grohs, Adam and Knill, 2015). Yet here, responses that contained a full answer were statistically indistinguishable from those that just contained contact information on perceptions of quality, friendliness, and helpfulness.

In both experiments, the number of days it took for the legislator to respond had a negative and statistically significant effect on perceptions of response quality and helpfulness. Perhaps unsurprisingly, the timeliness of the response had no significant effect on evaluations of response friendliness. Subjects were also asked to evaluate the speed at which responses were written to the constituent in order to gauge constituents’ views of “timeliness.” Figure 1 shows the relationship between the number of days until the legislator responded and perceptions of speed and quality, fitted using locally weighted smoothing (LOESS). Evaluations of response speed immediately decrease with each additional day. In both experiments, ratings of response speed decrease over 20 points within the first 10 days. However, in the first experiment, this does not affect respondents’ views on the quality of the response until about 2 weeks. In the second experiment, there are negative consequences for the perceived

response quality immediately. As respondents view the responses as less “timely”, they also view them as being lower in quality. Within the first 10 days, evaluations of overall response quality drop by over 15 points. The total drop in perceptions of response quality over the 30 day period is 10-15 points in both experiments.

Figure 1: Effect of Number of Days Until Response on Perceptions of Speed and Quality



Note: Fitted using locally weighted smoothing (LOESS). Grey shaded areas represent 95% confidence intervals. Solid line shows the effect of days on overall quality; dotted line shows effect of days on ratings of response speed.

3.2 MTurk Results

This section presents the findings from the MTurk study. Table 4 presents results from two OLS regression models estimating the effects of the email characteristics (coded by my research assistants) on the average level of satisfaction respondents rated each legislator response (recall satisfaction with response was registered on a scale from 0-100). Both models exclude respondents that failed the attention check, but the magnitude and statistical

significance of the coefficients do not change markedly when I include these respondents.⁵

Table 4: Effects of Response Characteristics on Satisfaction with Response

Automated	-22.616*	-20.912*
	(2.634)	(5.324)
Rep. (not staff)	2.664	2.140
	(1.474)	(1.471)
Named greeting	6.183*	4.860*
	(1.676)	(1.763)
Contact info	2.175	1.162
	(1.930)	(1.898)
Website	7.176*	5.334*
	(1.507)	(1.598)
Thank you	-1.432	0.250
	(1.535)	(1.536)
Ask more info	-3.539	-1.457
	(1.872)	(2.276)
Invite follow up	5.266*	6.128*
	(1.653)	(1.648)
Sign off	3.762*	3.329*
	(1.464)	(1.512)
Length	7.037*	5.008*
	(0.854)	(0.875)
Question answered		11.616*
		(1.911)
Constant	27.785*	20.911*
	(3.187)	(4.330)
Observations	400	313
R ²	0.557	0.553
Adjusted R ²	0.545	0.536

Note: *p<0.05; Standard errors are in parentheses. “Length” in Model 1 is measured as the natural log of the number of words to preserve all 400 observations; “Length” in Model 2 is measured as the natural log of the number of characters of each email and was taken from Butler’s (2014) replication data.

Model 1 includes all of the independent variables that were coded for the 400 emails. The response characteristic with the largest effect on satisfaction with the response is whether or not the email was an automated form message. Respondents were 22.6 points less satisfied

⁵18% of respondents failed the attention check.

with responses that were automated than those that were personalized. While it is possible that some of these automated responses were then followed up with a personalized response, it is often assumed in audit experiments that scholars use whatever email arrives first in response to the constituent request, even if it is an automated reply, and it is unclear how additional replies are counted, if at all.⁶

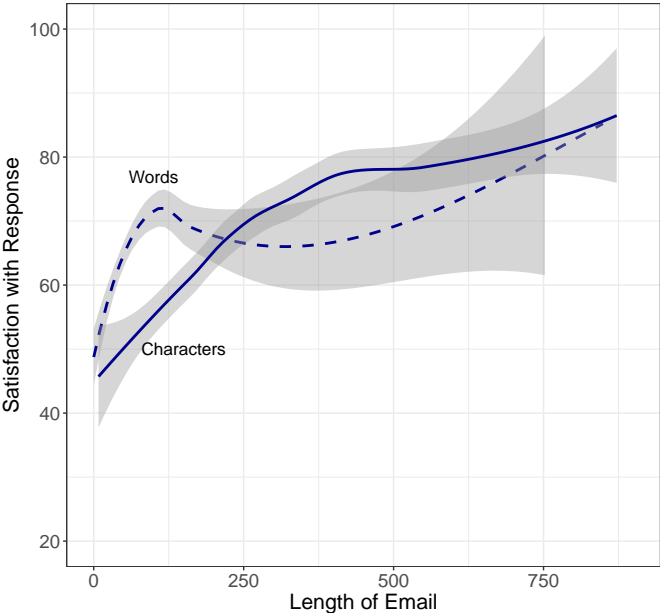
As for the friendliness items used in the survey experiments above, both named greetings (“Dear Rose”, “Hello Tremayne”, “Hi Mr. Joseph”, etc.) and invitations to follow up with additional questions separately had a positive and statistically significant effect on satisfaction with response. Additionally, when legislators ended their reply with a sign off (“Best”, “Sincerely”, “Regards”, etc.) respondents were 3.7 points more satisfied with the response than when legislators did not use a sign off. Including a link to a website also made respondents more satisfied with the response by 7.2 points.

Model 2 includes an additional variable that Butler coded for the full database of legislator responses in the audit experiment. This variable indicates whether or not the legislator answered the question asked by the putative constituent. With this variable included, the model includes 87 fewer observations because Butler’s replication data only includes the latest response sent by the legislator. For example, if the legislator’s office sent an automatic reply and then later followed up with a personalized response, only the latter email appears in the database. In this model, the coefficients for the other independent variables do not significantly differ in magnitude from those presented in Model 1. However, the variable for whether the constituent’s question was answered had a relatively large and statistically significant effect on respondents’ satisfaction with the response. Respondents were 11.6 points more satisfied with responses that answered constituents’ questions than responses that did not contain an answer.

⁶In this particular audit study, Butler (2014) did appear to replace automated messages if they were followed up with a subsequent email.

In both models, the length of the email also had a statistically significant effect on satisfaction with the response. I account for the length of the emails using words in the first model and characters in the second model. Character count was taken from Butler’s (2014) replication data so word count is utilized in the first model to preserve all 400 observations. These two measures are highly correlated ($r=0.96$) and I use the natural log of both variables since I expect there will be diminishing returns to increasing length; this can indeed be seen in Figure 2 which plots the bivariate relationship between length of email and satisfaction with response. The coefficients from the model in Table 4 indicate that a 10% increase in the word count results in a .7 higher satisfaction rating, whereas a 10% increase in the character count results in a .5 higher satisfaction rating.

Figure 2: Effect of Length of Email on Satisfaction with Response



Note: Fitted using locally weighted smoothing (LOESS). Grey shaded area represent 95% confidence intervals. Solid line shows the effect of characters on satisfaction of response; dotted line shows effect of words on satisfaction of response.

The coefficients that are not statistically significant in either model are just as telling as those that are. For example, asking for more information did not decrease respondents’

satisfaction with the response, even controlling for whether the constituent’s question was answered in the email. This is notable given the fact that some scholars consider emails like this to be non-responses. For example, McClendon (2015) notes that one of her measures of responsiveness “can be somewhat unfair to councillors who, in asking, for more information, were actually trying to be helpful and precise” (7). Similarly, only providing contact information for someone else did not necessarily decrease respondents’ satisfaction with the response, consistent with the findings from my experiments, and including a “thank you” did not improve satisfaction, although some scholars give extra credit to legislators that thank their constituents for writing (Carnes and Holbein, 2015). Moreover, the sender of the response had no statistically discernible affect on respondents’ satisfaction with the response; that is, whether the response was written by the legislator him or herself as opposed to a staff member mattered little to the rating of the response. However, some of this information was not obvious in the emails and it is likely that not all respondents were aware of the sender’s identity. Future work should more systematically look into whether constituents are less satisfied when a legislative staff member responds to their request rather than the legislator himself.

In sum, the findings from this study support what I found in the survey experiments and also provide new information regarding how individuals perceive legislative responsiveness. Respondents consistently preferred emails from elected officials that were friendly, whether that came in the form of a named salutation, sign off, invitation to follow up with more questions, or a personalized email rather than an automated one. Respondents also judged emails based on their length and timeliness. Finally, answering the question did have a statistically significant effect in the MTurk study, as well as proving a link to a website for the constituent to visit.

4 Weighted Measure of Responsiveness

A main goal of this paper is to increase the external validity of measures of responsiveness in audit experiments by examining how individuals perceive and evaluate communication from elected officials. I therefore use the results presented above to develop a new measure for elite responsiveness that can account for several facets of responsiveness at once. If we know what citizens deem a satisfying legislator response, we can weight responses by how well they meet those criteria.

The results from the previous three tests show that the most satisfying response from an elected official is one that is not automated, answers the question, arrives promptly, is at least 400 characters in length, and includes a named greeting, invitation to follow up, link to a website, and a sign-off. Responses that meet all these criteria are coded as 1. Following the intuition that constituents become less satisfied as one or more of these conditions is not met, I created a discount parameter that deducts from a full response (coded as 1) by the amount that condition is specifically worth. For this parameter, I used the coefficients from the previous sections to calculate the amount that should be discounted if a response lacks those particular characteristics. I divide the coefficients by 100 to reflect the fact that the weighted response measure is on 0 to 1 scale.

For example, for each additional day the legislator took to respond, I added 0.0033 to the discount parameter, which is based on the average of the coefficients from the models estimating overall quality in the first two experiments. I also cap this penalty at 30 days, so that after 30 days no additional penalty is incurred for responding “slow.” To determine how much weight to give to the length of responses, I ran the regression model from Table 3.2 without taking the log of characters, but instead capping the length at 400 characters, in order to estimate the linear relationship between the length of the email and respondents’ satisfaction. I implement the cap because the relationship between length and satisfaction

rating is linear up to 400 characters, as seen in Figure 2. Based on the estimated linear relationship, for every reduction in characters below 400, the response is penalized by .0005 points. Thus, for example, an email that is 300 characters in length would be discounted by .05 $((400-300) * .0005)$.

For the rest of the terms in the discount parameter, I used the coefficients from the full model in Table 4 to determine how much an email should be discounted if it is missing an important element. For example, an email lacking a named greeting would have 0.048 added to the discount and one without a website link would have 0.0533 added.

$$\begin{aligned}
 \mathbf{discount} &= \text{automated} * 0.209 \\
 &+ \text{no named greeting} * 0.048 \\
 &+ \text{no invite follow up} * 0.0613 \\
 &+ \text{did not answer question} * 0.116 \\
 &+ \text{no website link} * 0.0533 \\
 &+ \text{no sign-off} * 0.0333 \\
 &+ \text{characters} < 400 * 0.0005 \\
 &+ \text{days until response} * 0.0033
 \end{aligned}$$

I then created a new weighted response variable based on this value. I subtracted the discount from 1 for all responses received from legislators in the experiment. This weighted response can be used in place of the typical indicator variable for whether or not a legislator responded, included a full answer, was friendly, was timely, etc. By accounting for all of these factors of a response and incorporating it into a single measure, we can get a more parsimonious picture of how well legislators respond.

5 An Application

To demonstrate how this weighted dependent variable can be applied, I replicate the experiment in Butler (2014), which is the source of the raw emails used in the MTurk study. Since I (so far) only coded 400 of the 6,989 emails for the full set of independent variables described in described in Section 3.2, I cannot employ the full weighted replication on all of

Butler's data. I am only able to use the variables he coded for the entire database of emails. Specifically, I can weight the responses by whether they included an invitation to follow up, whether they answered the question, their length measured by number of characters, and after how many days the response was received. In a future iteration of this paper, I will code the remaining emails on the additional variables (named greeting, sign off, automated, etc.) so that I can use the fully weighted response measure detailed in the above section.

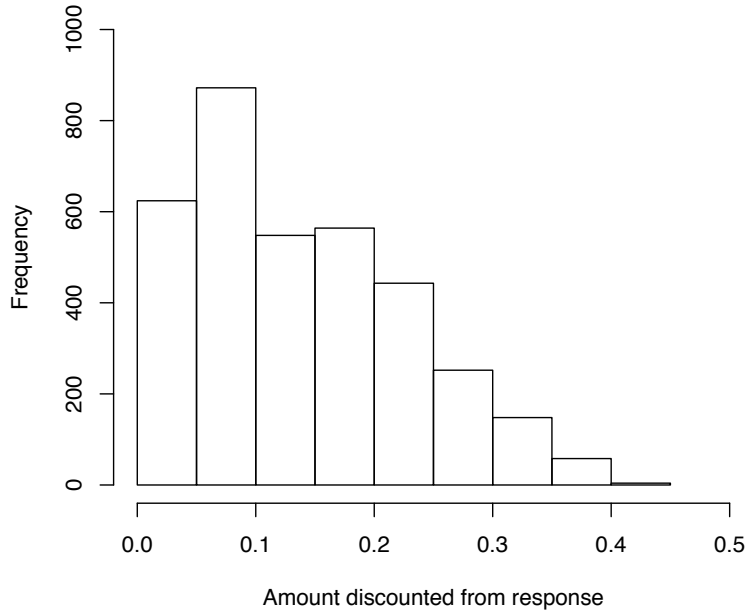
With regard to timing, Butler recorded the day and month emails were sent to legislators and the day and month a response was received. I created a variable based on these values for the number of days it took legislators to respond. 0 means a response was received the same day an email was sent out. Any value above 30 was set to equal 30 since that was the cut-off in the survey experiments reported in this paper.

Butler also coded whether or not the response included an invite to follow up via email or phone, but ultimately did not use this in his analysis. In line with the discount formula in the previous section, I add 0.0613 to the penalty when the legislator or staff member did not invite the constituent to follow up with additional questions.

Figure 3 shows the distribution of the discount as applied to Butler's full database of emails. Note that even the worst response only gets discounted by about 0.45. This suggests that even poor responses that perhaps did not require much effort on the legislator's part are still more worthwhile than not responding at all.

Table 5 presents the results from the exact replication of Butler's (2014) audit experiment as well as the preliminary replication using the weighted outcome measure of responsiveness. In his experiment, the response rate of non-Latino white legislators to the white alias was 6.4 points higher than to the black alias. The response rate of black legislators to the white alias, on the other hand, was 3 points lower than to the black alias. This is Butler's main treatment effect. Using the (partially) weighted measure of responsiveness, the gaps between the white and black alias' slightly diminish; the response rate of white legislators to whites

Figure 3: Distribution of discount variable



is only 5.5. points higher than to blacks, and the response rate of black legislators to whites is only 1.8 points lower than to blacks.

While the difference between his results and the weighted results are not statistically significant, the results suggest that while white and black legislators respond at different rates to white and black constituents, the quality of their responses towards both white and black constituents might not be as different as the overall rate of response. Future analyses will use the complete weighted measure to determine the full effect of accounting for the quality of responsiveness.

Table 5: Treatment Effect of White Alias on Legislator’s Responsiveness

	White Legislators		Black Legislators	
	Butler (2014)	Weighted	Butler (2014)	Weighted
Treatment effect	6.4*	5.5*	-3.0	-1.8
(Std. err.)	(2.7)	(2.4)	(6.9)	(6.0)

Note: *p<0.05

6 Suggestions for Future Audit Studies

In this section, I outline some concrete suggestions for future researchers measuring elite responsiveness to constituent contact. Aside from using the proposed weighted measure above, scholars can at a minimum account for several factors in their study to better measure the quality of legislator responses in an audit experiment.

1. **Be clear and intentional about measurement decisions.** At a minimum, it should be transparent to readers of any given study why some responses are considered “helpful”, for example, and others are not. This paper can be used as a guide for making these decisions, particularly in a pre-analysis plan to carefully outline a research design.
2. **Code for the friendliness of the response.** In all three tests, responses with “friendly” content, such as a named greeting up front, sign-off, and invitation to follow up with additional questions, were evaluated much more favorably than responses without these friendly items. Individuals perceived these emails not only as more friendly, but higher in quality and that they would be more satisfied with the response if they received it from their own legislator. This information can be used to weight the responses using the calculations presented in the section above, or responses with one or more of these items should be distinguished from responses without them.
3. **Track how long legislators take to respond.** In the two survey experiments, individuals started penalizing responses for being “slow” at different points. I therefore suggest following the weighting process in the above section and deducting 0.0033 each day that the legislator does not respond.
4. **Account for the length of the response.** Although only a few existing audit studies consider the length of emails when evaluating responsiveness, how many words or characters an email is can be a clear signal of the level of effort an legislator exerted

in responding to a constituent. Longer emails were rated higher than shorter emails, up to about 400 characters. For responses under 400 characters, I suggest deducting .0005 points for each character below that threshold.

5. **Be clear on how automated responses are treated.** Since impersonal, automated replies are over 20 points less satisfying to constituents than personalized replies, scholars should consider these responses separately. If this was the only response received from a legislator, I suggest discounting that response by 0.209 as opposed to treating it like a non-response.

References

- Berinsky, Adam J, Gregory A Huber and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political Analysis* 20(3):351–368.
- Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *The American Economic Review* 94(4):991–1013.
- Bishin, Benjamin and Thomas Hayes. 2016. "Do Elected Officials Service the Poor? A Field Experiment on the U.S. Congress." *Working paper* .
- Bol, Damien, Thomas Gschwend, Thomas Zittel and Steffen Zittlau. 2015. "The Electoral Sources of Good Political Representation A Field Experiment on German MPs." *Working paper* .
- 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.
- Burgess, Diana, Michelle Van Ryn, John Dovidio and Somnath Saha. 2007. "Reducing racial bias among health care providers: Lessons from social-cognitive psychology." *Journal of general internal medicine* 22(6):882–887.
- Butler, Daniel M. 2014. *Representing the Advantaged: How Politicians Reinforce Inequality*. Cambridge University Press.
- Butler, Daniel M. and Charles Crabtree. 2016. "Moving Beyond Measurement: Adapting Audit Studies to Test Bias-Reducing Interventions." *Forthcoming in Journal of Experimental Political Science* .
- Carnes, Nicholas and John Holbein. 2015. "Unequal Responsiveness in Constituent Services? Evidence from Casework Request Experiments in North Carolina." *Working paper* .
- Chen, Jidong, Jennifer Pan and Yiqing Xu. 2015. "Sources of Authoritarian Responsiveness: A Field Experiment in China." *American Journal of Political Science* .
- Costa, Mia. 2016. "How Responsive are Political Elites? A Meta-Analysis of Experiments on Public Officials." *Presented at the Southern Political Science Association Conference, January 2016* .
- Einstein, Katherine Levine and David M Glick. 2016. "Does Race Affect Access to Government Services?: An Experiment Exploring Street Level Bureaucrats and Access to Public Housing." *American Journal of Political Science* .
- Grohs, Stephan, Christian Adam and Christoph Knill. 2015. "Are Some Citizens More Equal than Others? Evidence from a Field Experiment." *Public Administration Review* .

- Jansson, Hanna and Per Adman. 2015. "Are Fateme and Abdelhakim being less informed by the government than Johanna and Martin? An experimental study of ethnic discrimination in Sweden." *Working paper* .
- McClendon, Gwyneth H. 2015. "Race and Responsiveness: An Experiment with South African Politicians." *Forthcoming in Journal of Experimental Political Science* .
- Pager, Devah. 2007. "The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future." *The Annals of the American Academy of Political and Social Science* 609(1):104–133.
- Vuolo, Mike, Christopher Uggen and Sarah Lageson. 2016. "Statistical power in experimental audit studies: Cautions and calculations for matched tests with nominal outcomes." *Sociological Methods & Research* 45(2):260–303.
- 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.

A Appendix

A.1 Definitions of “Good Response”

Table A1: “Good Response” Definitions

Study	Content	Tone	Timeliness
Broockman (2013)	<p>“I coded emails as helpful if they</p> <ul style="list-style-type: none"> •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 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>p. 2 of Supplementary Information</p>	NA	NA
Butler (2014)	<p>“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> <p>p. 30</p>	NA	2 weeks
Carnes and Holbein (2015)	<ul style="list-style-type: none"> •“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.]” •“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>p. 15-16</p>	<ul style="list-style-type: none"> •“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>p. 16 (alternate specification, but results not reported in paper)</p>	2 weeks

Chen, Pan and Xu (2015)	Deferral– reply but no answer; Referral– contact information for someone else; Direct Information– provides answer p. 11-12	NA	10 days
Einstein and Glick (2016)	NA	“whether the emailer is addressed by proper name. We were lenient in coding ‘yes.’ A named salutation could be as causal as ‘Hi Brett’ or as formal as ‘Dear Ms. Martinez’.” p. 13	24 hours
Grohs, Adam and Knill (2015)	<ul style="list-style-type: none"> •“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.” •“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>“We gave up to 3 points with regard to</p> <ul style="list-style-type: none"> •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 	15 days
Jansson and Adman (2015)	<ul style="list-style-type: none"> •“Number of questions answered: How many of the three questions are answered?” (03)” •“How much information does the public official provide in answer to the three questions? As little as possible (0) or broader information (1)? (range 03, the values for the answers to the three questions being added)” •“Does the public official give more information than asked for? (0 = no, 1 = yes)” p. 14 	<ul style="list-style-type: none"> •“Is the contact made more personal by using the sender’s first name? (0 = no, 1 = yes) ” •“Does the public official invite further contact? (0 = no, 1 = yes)” •“Is the sender welcomed to the municipality? (0 = no, 1 = yes)” p. 14 	NA

McClendon (2015)	<ul style="list-style-type: none"> •“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 	NA	NA
White, Nathan and Faller (2015)	“replies that provided links to state websites with official instructions about voting requirements.” p. 13 of Supplementary Information	“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	NA
Bishin and Hayes (2016)	<p>“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</p>	NA	NA

Note: Studies are listed in ascending chronological order. The according measurement and/or coding scheme are taken verbatim from the paper and are in quotations.

A.2 Treatment Emails

Answer

From: Representative Matt Johnson <MJohnson@malegislature.gov>
Subject: Re: Question on Voting



You must first register to vote. Once you are registered, you will have to provide an acceptable form of identification at the time of voting. I suggest checking our website (<https://malegislature.gov>) for more information about the voter registration process.

- Rep. Matt Johnson

Friendly Answer

From: Representative Matt Johnson <MJohnson@malegislature.gov>
Subject: Re: Question on Voting



Dear Jake,

You must first register to vote. Once you are registered, you will have to provide an acceptable form of identification at the time of voting. I suggest checking our website (<https://malegislature.gov>) for more information about the voter registration process.

Let me know if you have additional questions.

- Rep. Matt Johnson

Contact

From: Representative Matt Johnson <MJohnson@malegislature.gov>
Subject: Re: Question on Voting



I suggest contacting the elections office clerk. They can give you more accurate and up to date information about the voter registration process. You can email the office at atelections@ma.gov or call them at [\(413\) 555-7842](tel:413-555-7842).

- Rep. Matt Johnson

Friendly Contact

From: Representative Matt Johnson <MJohnson@malegislature.gov>
Subject: Re: Question on Voting



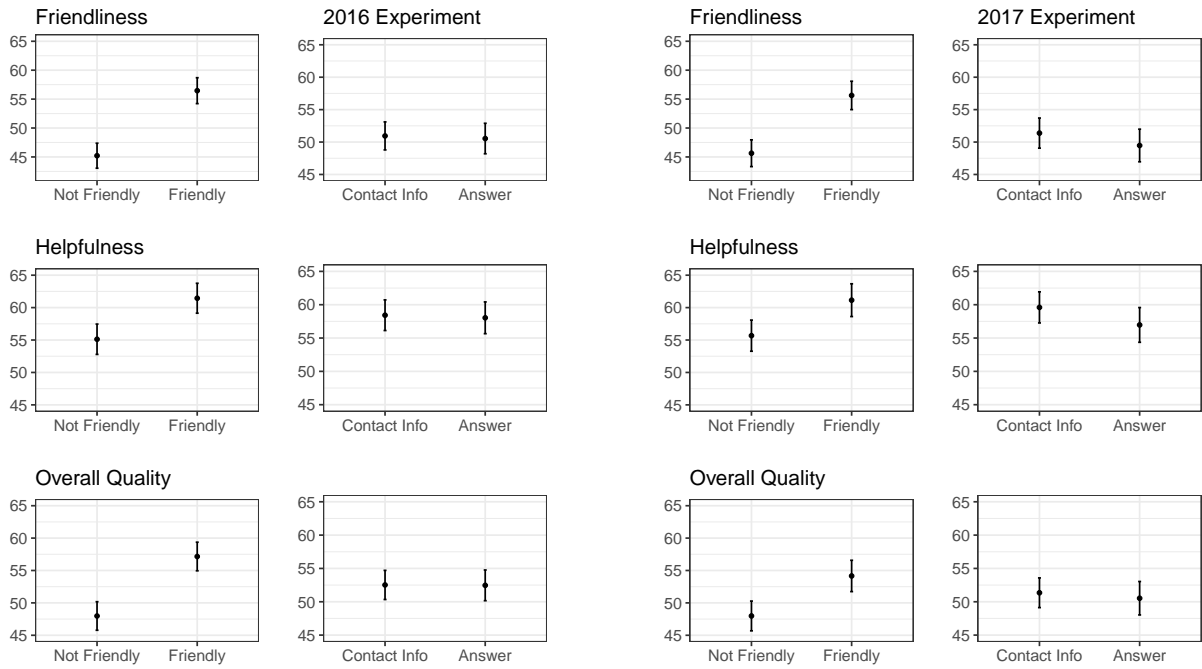
Dear Jake,

I suggest contacting the elections office clerk. They can give you more accurate and up to date information about the voter registration process. You can email the office at atelections@ma.gov or call them at [\(413\) 555-7842](tel:413-555-7842).

Let me know if you have additional questions.

- Rep. Matt Johnson

Figure A1: Mean Evaluations by Condition



Note: Plots show mean evaluations of respondents by condition, with vertical lines representing 95% confidence intervals. Two plots on the left show evaluations from the 2016 CCES experiment; the two plots on the right show evaluations from the 2017 experiment.