

Correspondence Audit Studies in Public Administration: A systematic review with commentary

By

Ryan A. Beasley and Joanna N. Lahey¹

Abstract: Correspondence audit studies are becoming more prevalent in the field of Public Administration (PA). We explain the benefits and limitations of audit studies and provide a systematic literature review of PA audit studies. Our systematic review includes journal articles written in English and containing audit studies in PA that were gathered from highly ranked PA journals, and also forward and reverse citations of relevant prominent articles. We summarize the frequency of topics and other selected aspects of the 90 included articles. We then discuss common concerns with audit studies both in PA and more generally. Finally, we discuss future directions for these kinds of studies that go beyond the standard measures of discrimination between two groups.

Acknowledgements: Thanks to Jacqueline Mendez, Lala Taghiyeva, Zakary Piwetz, Brady Duke, Ngoc-Nhu Vo, Karen Gracia, and Gabriela Bazan for excellent research assistance.

¹ Corresponding Author: jlahey@tamu.edu. Public Service and Administration, Texas A&M University. Ryan Beasley: dr.ryan.a.beasley@gmail.com, OncAI.

I. Introduction

An audit study, sometimes called a correspondence audit or correspondence review, is a form of field experiment in which the researcher sends out treatment and control stimuli, often written, to real participants who do not realize they are being studied. The researcher then measures differences in responses for the different treatments. Audit studies were pioneered in the mid-20th century to study discrimination in housing and labor markets (Fix & Struyk 1993). In these original studies, researchers followed up correspondence with pairs of in-person actors chosen and trained to be as identical as possible except for the group characteristic being studied, for example, race. Today in-person audits are rare, but correspondence audits have been growing in popularity across many disciplines (Gaddis 2018). Modern audits utilize new techniques to better study the market, such as investigating interaction effects between stimuli characteristics and presenting stimuli via a variety of digital media. Given the large size of many modern audits, researchers are less likely to send pairs of stimuli which are identical except the item being studied. Additionally, while early audit studies usually focused on race discrimination, modern researchers use audits to study any outcome that involves differential treatment.

This emphasis on hypothetical stimuli sent to real participants who 1. make real decisions, and 2. do not know that they are being studied, is what sets audit studies apart from other A/B style field experiments. Audit studies are not surveys or survey experiments in which the participants know their responses are being studied. They do not involve providing vignettes and asking for hypothetical choices. They do not use a student sample as a substitute for non-student decision-makers. The benefit to these restrictions is that well designed audits provide information about how actual constituents respond in real market environments. That said, it can be difficult to determine whether certain field experiments should be classified as correspondence or audit studies. For the purposes of this review, we are defining audit studies as studies that attempt to *measure* behavior, not to *change* behavior. Examples of attempting to change behavior (thus are not audit studies) include experiments that nudge taxpayers into paying real tax bills (John & Blume 2018), vary advertising messages to recruit police applicants (Linoss 2018), or change stimuli presented in Facebook ads to see how they can affect future polls (Broockman & Green 2014).

The primary benefit of audit studies is that, unlike cross-sectional studies using large datasets, they can measure behavior at the partial equilibrium level rather than the total equilibrium. In other words, audit studies can measure discrimination as it occurs at the application level between otherwise identical applicants. The final outcome of any market includes selection, in which people avoid applying to places in which they believe there is discrimination, or in which people who are discriminated against must spend additional effort compared to those who are not; thus studies that look only at the final outcome may miss discriminatory behavior. Additionally, in many markets, the average characteristics of the different groups of applicants are not identical, so without an audit study, it is unclear whether general equilibrium differences in outcomes are caused by discrimination or by differences in the applicants themselves. Audit studies bypass both of these problems.

While the audit study is a powerful research method, it has limitations. First, audit studies generally do not measure the last step in a multi-stage process. For example, if measuring discrimination in labor markets, they only measure whether a hypothetical applicant is called in for an interview, not if the applicant receives a job offer. If discrimination increases or decreases at later stages in the process, audit studies are unlikely to pick up these differences. However, treatment at this first stage is important by itself: one cannot get to the second stage without passing the first stage. Many PA studies do not have this limitation because they focus on one-stage processes such as one-time requests for information to government officials. However, even these studies have the problem that without being able to see the entire market: it is unclear whether, for example, Republicans tend to be more responsive than Democrats to constituent stimuli (Hayes & Bishin 2020; Janusz & Lajevardi 2016) because they have more responsive natures or because they get fewer requests.

Another limitation of the audit study is that they are only externally valid for the stimuli that they send. We have coined the term “template bias” to describe one aspect of this problem (Lahey & Beasley 2009). Template bias occurs when a limited number of stimulus templates are sent out that do not represent sufficient variation. For example, assume a hypothetical labor market in which Hispanic men are only discriminated against if they have mustaches, while non-Hispanic men with mustaches are not discriminated against. If only auditors without mustaches are sent to apply for jobs, this underlying discrimination will not be found. Ideally an audit study will send stimuli that fully explore the feature variation seen in the market and relevant to the study topic.

Finally, audit studies tend to be limited to brief interactions and limited outcome variables; they generally measure the outcomes of the first stage of an application process or a brief request for information. Unlike a survey, focus group, or other more in-depth approach, audit studies cannot collect information on counterfactuals and are generally limited to what demographic information is publicly available. There are some exceptions that combine an audit study with another form of experimental design, but these are rare (Dynes et al. 2021; Rooth 2010). Similarly, audit studies are unable to include all possible participant characteristics; for external validity they can only present stimuli that could be seen in the market.

The audit study method is gaining in popularity for gathering evidence related to Public Administration (PA) hypotheses and models, and while systematic reviews have been done on the more general domain of PA field experiments (Bouwman & Grimmelikhuijsen 2016; Hansen & Tummers 2020; Li & Ryzin 2017), this is the first effort to systematically review audit studies and correspondence reviews in PA (Public Administration, Public Policy, and relevant Political Science academic journals). By systematically reviewing these studies, we hope to provide an overview of the innovations that PA has brought to audit studies (in outcomes, inputs, and participants), as well as a starting point for new practitioners.² We refer readers interested in how to do an audit study to the Lahey & Beasley (2018) chapter in Gaddis (2018).

² There are many purposes of systematic literature reviews. We are using the first listed purpose of a systematic literature review from (Page et al. 2020, p. 1), to “provide syntheses of the state of knowledge in a field, from which future research priorities can be identified.” We are hoping to provide information to PA researchers about

II. Systematic Literature Review Methods

The eligibility criteria for this systematic literature review are Public Administration research articles published in peer-reviewed academic journals. Only English articles were included. Any publication date was eligible, with new papers gathered through May 2021.³ The only eligible study designs were audit studies or correspondence reviews in which the participants performed a real task, i.e., a typical action they might perform such as one of the expected tasks of their job, and were unaware of their study participation while doing so. The participants, tasks, and/or theoretical framework must have been relevant to Public Administration.

Articles were gathered from specified journals, papers cited by selected systematic literature reviews, and forward and backward citations. We first searched the top 50 journals ordered by Scimago Journal Rank (SJR) indicator, from the Scimago Journal & Country Rank website under subject area "Public Administration" [<https://www.scimagojr.com/journalrank.php?category=3321> last accessed 5/13/21]. We searched each journal via its website for ("audit" AND "Experiment") OR ("Correspondence" AND "Experiment"). Additionally, we gathered papers that cite, or are cited by, three systematic literature reviews on experimental studies in PA (Bouwman & Grimmelikhuijsen, 2016; Hansen & Tummers 2020; Li & Ryzin, 2017). We also gathered relevant papers that cite, or are cited by, papers collected from the above sources that had particularly impressive literature reviews or seemed to be cited by many papers (Adman & Jansson 2017; Broockman 2013; Costa 2017; Landgrave 2020; Margetts 2011).

As the papers were collected, they went through an exclusion screen to remove papers that were 1. Not an experiment, 2. A survey, vignette, or other hypothetical experiment in which participants know they are being studied, 3. Not Public Administration, and 4. Trying to change behavior (beyond any immediate response to the stimuli) rather than just measuring it. Two reviewers, one PhD coauthor (Beasley) and one graduate research assistant (Jaqueline Mendez), worked independently to determine which studies met the inclusion criteria. Reviewers read the title and abstract and when the abstract did not provide enough information, they read the paper itself. Questions and disagreements were handled by discussion with the other PhD coauthor (Lahey). On a regular basis an additional graduate research assistant (Zachary Piwetz) compared the lists from the two reviewers, and the team then met to discuss any discrepancies between the lists.

Several RAs worked independently to pull information out of the papers for summary statistics. Each included paper had two to three research assistants separately screening for characteristics. Any discrepancies were brought to the attention of the authors who made final decisions on any disagreements. Table 1**Error! Reference source not found.** contains the aspects that were coded for

the field of audit studies within the broader area of public administration (in which we also include politicians and non-profits).

³ Some papers in the review will have later final publication dates because they were cited while working papers or they were published online prior to a later publication date in the paper version of the journal. The following unpublished papers would have been included if they had been published: De Vries, Dinas, & Solaz 2016; Janusz & Lajevardi 2016; Mendez 2014; Mikula & Montag 2022; and Timm 2021.

each article broken apart by participant type. With some important exceptions, it appears that there are more similarities than differences between studies of street-level bureaucrats and elites, so we will generally be discussing studies with all types of participants together, pointing out some important differences in the discussion.

Table 1. Study design features

Study design features	Participants			Total
	Street-level bureaucrats	Political Elites	Other	
<u>Participant</u>				
Ideological families or political parties	4	23	4	31
Gender	1	14	2	17
Race or ethnicity	2	8	5	15
Job security	0	10	0	10
Country	1	2	2	5
Pre-election vs post-election	0	2	2	4
Methods of election (e.g., nominal vs party list)	0	4	0	4
Status of being impacted by political mandates or antidiscrimination policies	1	0	1	2
<u>Stimuli</u>				
Race or ethnicity	14	21	10	45
Gender	7	16	2	25
Tone or consequences of request (including FOIA)	9	7	3	19
Socio-economic status	5	7	3	15
Political party	2	7	2	11
Constituency	1	8	1	10
Personal vote intentions	0	6	0	6
Religion	2	3	0	5
Social or peer pressure	3	2	0	5
Incentivization of participation	1	2	1	4
Citizenship	0	2	1	3
Policy requests vs service requests	0	2	1	3
Inclusion of information about policies in other countries/states/areas	1	1	1	3
Political donor statement	0	2	0	2
Same-sex partner	1	0	0	1
<u>Analysis</u>				
Between-subjects	26	36	15	77
Standard OLS/logit/probit regression	23	31	17	71
Evaluates the quality or speed of responses	23	26	12	61
Interaction effects (e.g., writer's race vs the participant's party)	9	14	4	27
Within-subjects	3	7	3	13
Includes a power analysis	3	4	1	8
Effect of an initial stimulus on a later stimulus	1	1	0	2

Note: Other includes both elected officials and street-level bureaucrats in the same study (4), employers (3), universities (2), potential voters (1), nursing homes (1), salespeople (1), landlords (1), facebook users (1), citizens (1), and other private/non-profit services (3). (If the participant was not a public official, the stimulus was related to public affairs writ large.)

III. Systematic Literature Review Results

Our systematic literature review identified a total of 2076 references: 1282 through the journal search, 255 from the previous systematic literature reviews, and 539 through cited or citing reference searches, as shown in Figure 1. We excluded 1986 references. A total of 90 journal articles are included in the results below. As noted earlier, we excluded Broockman & Green (2014), John & Blume (2018), and Linos (2018) because they attempted to change behavior rather than just measure it. We included Michelson (2006), on the other hand, because some of the analysis is on whether canvassers were able

to contact participants, even though the rest of the study (whether the participants voted at a later date) was about changing behavior. Similarly, we included Butler & Crabtree (2017) even though it measured the behavioral impact of receiving an information treatment at an earlier date, because two-thirds of the participants did not receive that information and results were reported separately by group. We included Jilke et al. (2019) because it manipulated information cues to test immediate behavior, not to influence future behavior. For a tricky example of participants knowing they were in an experiment, we excluded Nyhan & Reifler (2015) because the participants in the placebo group were sent a letter informing them about the experiment and their results combine the placebo and treatment groups. We were generous in our inclusion of which papers were of interest to PA, including papers from our list of journals such as Darolia et al. (2015), published in the *Journal of Policy Analysis and Management*, which studies whether employers prefer workers from for-profit vs. not-for-profit colleges. We also included papers studying politicians. We excluded Schultz, Khazian, & Zaleski (2008), which was cited in a prominent paper, but was published in a psychology journal, because its topic of promoting conservation among hotel guests is not PA. Our exclusion of non-audit papers also meant that we excluded papers that analyze already gathered data, e.g., Lowande, Ritchie, & Lauterbach (2019).

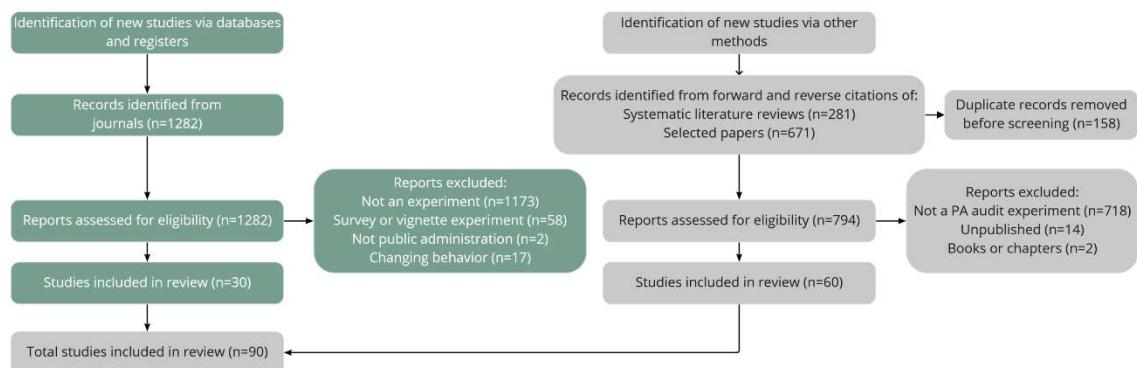


Figure 1. Consort diagram

In the past two decades, 90 PA journal articles that use audit studies have been published, with the rate of publication rapidly increasing over the past decade, shown in Figure 2. The studies were gathered from 47 journals, 15 of which published more than one PA audit study, shown in Table 2. These studies were performed in many countries: Argentina, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, Denmark, England, Estonia, Finland, France, Germany, Hungary, India, Ireland, Italy, Mexico, Netherlands, New Zealand, Russia, Slovakia, South Africa, Sweden, Switzerland, Taiwan, United Kingdom, United States of America, and Uruguay.

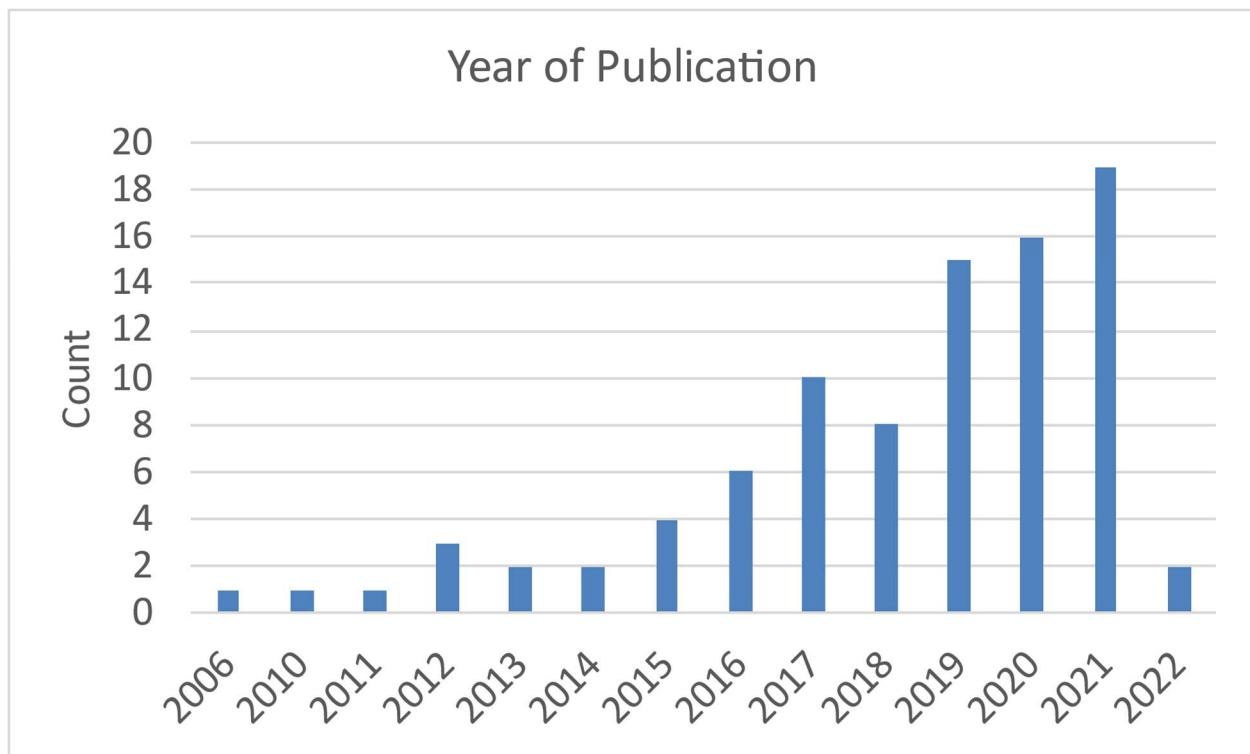


Figure 2. Year of Publication

Table 2. Journal of Publication

Journal		Participants			
		Street-level bureaucrats	Political Elites	Other	Total
American Journal of Political Science		5	6	2	13
Journal of Politics		1	4	2	7
Journal of Experimental Political Science		1	5		6
Journal of Public Administration Research and Theory		2		3	5
Public Administration Review		4			4
Journal of Behavioral Public Administration		2	1		3
Government Information Quarterly		2		1	3
Legislative Studies Quarterly			3		3
Governance		2			2
American Political Science Review		1	1		2
Research & Politics		1	1		2
Social Science Quarterly		1		1	2
Political Research Quarterly			2		2
American Politics Research			2		2
European Journal of Political Research			1	1	2
Total		22	26	10	58

Notes: Journals with exactly one included publication: Administrative Science Quarterly; British Journal of Political Science; China Quarterly; Cogent Social Sciences; Criminology & Public Policy; Democratization; Indian Review; International Interactions; Journal of Civil Society; Journal of Comparative Economics; Journal of Law and Economics; Journal of Policy Analysis and Management, Journal of Race, Ethnicity, and Politics; Journal of the European Economic Association; Local Government Studies; New Media and Society; Parliamentary Affairs; Party Politics; Perspectives on Politics; PLOS ONE; Political Analysis; Political Psychology; Political Science; Political Science Research & Methods; Politics & Gender; Politics, Groups, and Identities; Proceedings of the National Academy of Sciences; Public Administration; Quarterly Journal of Political Science; Regulation & Governance; Social Science Journal; State Politics & Policy Quarterly. See Table 1 for additional notes.

Who is studied and How?

A distinguishing feature of PA audit studies is the typical participant pool: elected and appointed officials or their constituents. Some studies include a wide variety of officials, but many focus on a specific subset, such as politicians, front-line street-level bureaucrats, or back-office administrators. This participant pool differs from those used in economics and sociology which are more likely to focus on labor, mortgage and rental markets (see Gaddis 2018 for a literature review), or the emerging public health literature on customer service (e.g., Button et al. 2020; Mackenzie-Liu et al. 2020). Overall, 43 of these studies focus on political elites, 29 on bureaucrats, and 18 have participants who do not belong to one of these groups (but some other aspect of the study makes it of interest to public administration researchers), as shown in Table 3.

Table 3 also presents the different outcomes studied in PA correspondence studies. The most popular topic for these studies in PA (43% of included papers) is the responsiveness of political elites to unofficial requests for information. Another popular area of study instead investigates the responsiveness of street-level and frontline bureaucrats to different constituent characteristics (23%, e.g., Adman & Jansson 2017; Distelhorst & Hou 2014; Einstein & Glick 2017; Grohs et al. 2016; Michener et al. 2020). Nearly as popular (14%) is another type of study in which researchers present freedom-of-information act (FOIA) requests varying tone, information provided, and writer characteristics such as political donations or demographic characteristics (e.g., Ben-Aaron et al. 2017; Jenkins et al. 2020; Lagunes & Pocasangre 2019; Michener et al. 2020; Peisakhin 2012; Spáč et al. 2018). While some papers study job callbacks (Darolia et al. 2015; Hou et al. 2020; Kang et al. 2016; Vuolo et al. 2017), they are uncommon (4%), particularly compared to audit studies done by sociologists and economists (see Gaddis 2018 for a literature review and Lippens et al. 2023 for a meta-analysis)⁴.

⁴ Other literature reviews of field experiments more generally include Bertrand and Duflo (2017), Neumark (2018). Meta-analyses on audit studies using racial discrimination specifically include Quillian et al (2017) and Gaddis et al. (2021).

Table 3. Outcomes of interest

Outcome	Participants				Total
	Street-level bureaucrats	Political Elites	Other		
Responsiveness to unofficial request	21	39	6	66	
Responsiveness to FOIA request	7	3	3	13	
Callbacks to job applications			4	4	
Other	1	1	5	7	
Total	29	43	18	90	

Note: Other Outcome includes voter turnout, ID declination, charitable donations, willingness to participate in a survey, rental approvals, and offers of investment incentives. FOIA row includes other (non-US/state) official freedom of information/public records act requests. See Table 1 for additional notes.

PA audit studies often explore how participant responses vary with respect to participant characteristics, for example, do participants from different demographic or political groups respond differently to stimuli. Many experiments have analyzed differences with respect to the participant's race and ethnicity (17%, Broockman 2013; Butler & Broockman 2011; Gaikwad & Nellis 2021; Habel & Birch 2019; McClendon 2016; Mendez & Grose 2018; Newland & Liu 2021; Vaishnav, Khosla, Milliff, & Osnos 2019) or the participant's gender (19%, Dhima 2022; Magni & Ponce de Leon 2021; Rhinehart 2020; Thomsen & Sanders 2020). Responses have been compared between participants from different ideological families or political parties (34%, Butler, Karpowitz, & Pope 2012; Dropp & Peskowitz 2012; Gell-Redman et al. 2018; Hayes & Bishin 2020; Porter & Rogowski 2018). The response rates of legislators (political elites) have been examined as a function of their job security (11%, Dropp & Peskowitz 2012; Giger et al. 2020), and pre-election vs post-election timing (4%, Driscoll, Cepaluni, Guimarães, & Spada 2018). Similarly, the behaviors of nominally-elected legislators have been compared with those elected via party list (4%, Bol 2021). The hiring practices of firms impacted by political mandates have been compared with firms free of such mandates (Hou, Liu, & Crabtree 2020), and the responsiveness of participants affected by antidiscrimination policies have been compared with participants not affected by such (2%, White, Nathan, & Faller 2015). Responses by local municipal officials have been compared across countries (Butler, De Vries, & Solaz 2019). Incorporating participant characteristics as these studies do enables testing of more nuanced models of behavior, often detecting heterogeneous behavior between subpopulations of participants, and occasionally showing that all the detected discrimination from the participants is coming from such a subpopulation (e.g. Mendez & Grose 2018).

Stimuli signals

As is common in other fields, many PA audit studies vary the stimuli to signal differences in the hypothetical writer's race or ethnicity (50%, Butler & Broockman 2011; Landgrave 2021; Newland & Liu 2021; Olsen et al. 2022), gender (28%, Golder et al. 2019; Kalla et al. 2018; Magni & Ponce de Leon 2021), socio-economic status (17%, Hayes & Bishin 2020; Taghizadeh et al. 2022), and religion (6%, Crawfurd & Ramli 2022; Lajevardi 2020; Pfaff et al. 2021; Vaishnav et al. 2019). For the most part, studies on ethnicity, gender, immigrants, race, and religion find discrimination against females and minorities, both in response rate and response quality. However, the subset of studies in which stimuli include political aspirants asking for guidance generally present pro-women results (e.g., Rhinehart 2020; Dhima 2020). Additionally, a small number of non-US studies do not find discrimination (e.g., Newland & Liu 2021; Taghizadeh et al. 2022), or in one case find pro-women results (Magni & Ponce de Leon 2021). Discrimination can be reduced by suggesting that the participant's responses may be publicized (Landgrave 2020), or by contacting political candidates in the time leading up to elections (Driscoll et al. 2018). Like other discrimination studies, studies on schools found religious/political/racial discrimination by officials, though Block et al. (2021) shows that racial discrimination found in higher education is less than what is found in the general public. In comparison, studies investigating SES and education have provided mixed results. Some show that stimuli indicating lower class/education get fewer responses (Landgrave & Weller 2022; Michener et al. 2020; Neiman 2017; Peisakhin 2012; Taghizadeh 2021), others show no evidence of bias (Carnes & Holbein 2019; Hayes & Bishin 2020; Lajevardi 2020). Unlike audits in labor economics, age-based signals are not yet a focus of studies in PA.

Other PA studies vary their stimuli in more novel ways.⁵ Studies find no effect of citizenship status when SES is controlled for (Alizade et al. 2021; Landgrave & Weller 2022; Newland & Liu 2021). They find that political elites are more helpful for constituents compared to non-constituents (Loewen & MacKenzie 2019). Senior policy makers made themselves available 3-4 times more often for political donors (Kalla & Broockman 2016). Legislators are more likely to respond when the stimuli includes an intention for personal voting instead of partisan voting (Bol et al. 2021). Finally, bureaucrats provisioning marriage licenses show no evidence of discrimination against same-sex partners (Lowande & Proctor 2020).

Beyond the hypothetical writer's characteristics, PA studies have also varied stimuli characteristics. For example, offices are more responsive to service requests than to requests about policy (Butler, Karpowitz, & Pope 2012). Compliance is increased by peer effects, but unaffected or reduced by shaming (Ben-Aaron et al. 2017; Terechshenko et al. 2019). When provided with information about implemented policies, local European officials are not biased against foreign policy information (Butler et al. 2019). Finally, legislators asked to take a survey and incentivized by a donation to a charity for college students are less likely to participate if the donation is for Hispanic students than for students of unspecified ethnicity, but the discrimination is mitigated if the stimulus says the donation will be made public knowledge (Landgrave 2020).

⁵ It is not our intention to discourage researchers from using standard stimuli. There is still much to be learned from standard stimuli.

A growing subfield of PA experiments investigates how officials respond to the “tone” of a stimulus or to potential consequences (21%, Chen, Pan, & Xu 2016). Those stimuli often include official requests under freedom-of-information laws (14%, Lagunes & Pocasangre 2019; Michener et al. 2020; Rodríguez & Rossel 2018; Spáč et al. 2018; Worthy et al. 2017). Peisakhin (2012) even compared requests made under such laws with requests accompanied by bribes. Most of these studies find that freedom-of-information laws are beneficial, with different studies looking into different aspects leading to responsiveness; for example, larger municipalities are more responsive (Spáč et al. 2018), the strongest effect is on concordance (not compliance) (Grimmelikhuijsen et al. 2019), stating the law is only for male requestors (Rodríguez & Rossel 2018), and formal requests almost erase class differences (Peisakhin 2012).

Beyond the effects of either stimulus or participant characteristics on outcomes of interest, interactions between those characteristics can have important effects on outcomes. 30% of included papers analyze interaction effects. Discrimination is found in the interaction between ethnicity and either occupation (Habel & Birch 2019) or gender (Grohs et al. 2016). PA audits have studied in-group vs out-group association for ethnicity with local officials less likely to assist outgroup members (Distelhorst & Hou 2014), race with politicians more responsive to same-race constituents (McClendon 2016), gender with legislators more responsive to female political aspirants (Dhima 2020; Rhinehart 2020), and political party where political party of stimulus does not mitigate racial bias exhibited by political elites, while local election clerks are more responsive to co-partisan stimuli (Butler & Broockman 2011; Porter & Rogowski 2018). The intrinsic vs extrinsic motivation of the participant has been investigated via interaction effects between constituency and electoral safety—responsiveness to constituents is even higher with electoral competition (Giger et al. 2020), and between constituency and the participant’s race with Black legislators more motivated than non-Black legislators to respond to Black non-constituents (Broockman 2013). Discrimination against stimuli signaling different races and ethnicities has been studied via various interaction effects, for example, whether discrimination against Hispanics is driven by Republican legislators (Gell-Redman et al. 2018), participants supporting voter identification laws are less responsive to Latino stimuli (Mendez & Grose 2018), and bias against Latinos is lower in jurisdictions covered by antidiscrimination policies (White, Nathan, & Faller 2015). These studies show that interactions between characteristics can be necessary factors in understanding experimental results.

Study design and analysis

Along with identifying the participant pool and crafting stimuli content, two important study design questions are how many stimuli to send to each participant, and which medium to use for communication. With any within-subject design, a participant may directly compare stimuli, leading to them treating the stimuli differently than they would have in the absence of the comparison. Indeed, there is direct evidence of spillover effects within audit studies with multiple stimuli sent to each participant (Phillips 2019). In a field experiment such as a correspondence audit, there may be additional concerns about the participant realizing that they in an experiment should they receive multiple similar stimuli, and multiple stimuli mean that the participant may be more likely to treat the hypothetical information as real information, which could affect their beliefs and decisions. The majority

of included papers use between-subjects studies (86%); 81% of the papers sent a single stimulus to each participant. The stimuli medium can also result in different levels of responsiveness and response quality (Epstein et al. 2021). The majority of PA audit studies (88%) send stimuli via email due to its ease and minimal expense. Notable alternatives include in-person audits (7%; Ernst, Nguyen, & Taylor 2013), physical letters (6%; Butler, Karpowitz, & Pope 2012), and online forum posts or other website use (7%; Chen et al. 2016; Jilke et al. 2019). Phone calls are uncommon (3%). 10% of the studies incorporate multiple communication media.

Many studies (69%) include multiple outcomes of interest, with (68%) analyzing response rates plus various measures of the quality of response (Hemker & Rink 2017; White et al. 2015; Worthy et al. 2017). See Costa (2020) for a study on what alters the quality of responses, and Coppock (2019) for a short discussion on how to handle non-responses and avoid post-treatment bias. Only 9% of papers include a power analysis to show the smallest effect size that could have been detected (Landgrave 2020).

Innovations

PA as a field has brought broad innovations to the audit literature. Rather than focusing on labor and housing markets, PA has leveraged this technology to investigate important questions about the responsiveness of public servants, including responsiveness to FOIA requests. Another broad innovation stems from PA audits often knowing participant characteristics (particularly in the case of political elites), unlike many economics and sociology audits where an employment firm or housing address may be known, but the actual person making the decisions is unknown. That means that not only can PA audits determine differences in treatment by participant characteristics, but they can also study in-group vs. out-group differences and match participant and stimulus on demographic characteristics.

In addition to the general innovations of PA audit studies, during our systematic literature review we came across PA papers that we found to be particularly innovative. Multiple studies have been designed to accomplish two tasks simultaneously. Similarly to some FOIA audits (e.g., Jenkins et al. 2020; Spáč et al. 2018), Grose et al. (2015) directly used the responses to the stimuli in a follow-on study, by surveying how constituents reacted to actual legislator messaging. Both Block et al. (2021) and Landgrave (2020) did audit studies to see who would participate in surveys, but Landgrave additionally used the parameters of the survey incentivization (the specific recipient of the donation and the donation publicity/anonymity) as independent variables in the audit study. Vuolo et al. (2017) combined a job application audit study with an analysis of the calls for applications. Peisakhin & Pinto (2010) provided social benefit with their audit study by recruiting confederates that might materially benefit from participation.

Some included PA audit studies innovated study design and analysis. Innovative outcome variables include Butler & Crabtree (2017) and Butler et al. (2019) which study whether emailed stimuli were opened, something Hughes et al. (2020) claims is a metric of implicit discrimination. Butler et al. (2019) further tracked whether participants followed emailed links to additional information. As another

example, Lagunes & Pocasangre (2019) repeated the same FOIA audit study over multiple years and found changes in responsiveness over time.

More than one study tested innovative interactions. For example, Giger et al. (2020) and Broockman (2013) tested intrinsic vs extrinsic motivation in political elites by varying the constituency or jurisdiction conveyed by the stimuli. Michener et al. (2020) discovered that response rate depended on whether the putative stimuli writers were discoverable via websearch. Taghizadeh (2021) interacted ethnic signals in the stimuli with participants' caseloads to investigate the effects of workload on discrimination. Similarly, Habel & Birch (2019) found an interaction effect between occupation and ethnicity. Block et al. (2021) included officials and members of the general population, enabling comparison of their discrimination in responding.

In another example of innovation, Epstein et al. (2021) compared municipality responses by stimulus medium (email vs Twitter) and found differences in responsiveness and quality, making the point that studies will have to adapt as new media grow in public usage.

IV. Additional concerns

In this section, we provide additional considerations not mentioned above. Our purpose is not to confront the PA literature, but instead to provide areas for PA researchers to think about.

Although many audit studies do an excellent job of beginning from a theoretical or policy-relevant framework, as James et al. (2017) note in their book chapter on public management field experiments generally, some studies do not. James et al. (2017) also note that formal theoretical foundations are not always necessary and that results from field experiments can be used to formulate new theory. However, experiments are best when they are well thought out and are guided by at least informal theoretical considerations.

It is important for authors to do and present preliminary power calculations prior to sending out a study. This need is particularly important when the study finds null results, as insignificant results could be caused by insufficient sample size. In addition, audit studies that send multiple treatments to participants need to be sure to adjust their sample size calculations for clustering, as unadjusted size sample size calculations will provide too small a sample (see Lahey & Beasley 2018). And, as with all experiments, power calculations that intend to include interaction terms may need to more than double the single term sample size to get the correct power (see Giner-Soralla, 2018 for an excellent discussion of heuristics for interaction terms). It is concerning that less than 10% of the audit studies we found in the systematic review mentioned these calculations.

Two sources of internal validity concerns must be addressed when multiple stimuli are sent to the same participant. First, naively sending similar stimuli to a broader collection of participants could result in participants getting unusual stimuli (ex. correspondence from Black constituents when such correspondence is rare) that causes them to either suspect they are in an experiment or to form false beliefs about their constituent needs. A second internal validity concern arises when a participant compares stimuli because such comparisons can alter responses. Specifically direct comparison of

nearly identical stimuli can exacerbate difference in outcomes because when stimuli are nearly identical a small difference will drive the participant’s choice, or reduce them because a direct comparison makes the socially desirable choice more salient. A way to address these concerns is to limit the study to participants who receive many similar correspondences so the stimuli are just two (or more) of many and are thus less likely to be directly compared, although this could limit the external validity of the study. To further mitigate concerns about multiple correspondence, all stimuli features can be varied within reasonable limits (i.e., change more than just the name of the putative stimuli writer), when possible send stimuli that are not directly comparable (e.g., FOIA request vs simple ask about unrelated issue), and send stimuli at separate times.

A common problem with audit studies across all fields, not just in PA, is when researchers do not ensure their stimuli are externally valid. In some cases, they end up testing not what they think they are testing, but testing the unusual. Examples from economics might be showing pictures on resumes in countries where such things are uncommon, or omitting education information when education is always shown on resumes. More subtle versions of this problem occur when, for example, using names to signal race, gender, or ethnicity, because names also signal age, socioeconomic status, and potentially citizenship status (Einstein & Glick 2017; Barlow & Lahey 2018; Gaddis 2017a, b; Landgrave 2021; Crabtree et al. 2023). Similarly, a potential concern related to the external validity of the Y variable is pointed out by Olsen et al. (2022)—just as the request for an interview is only the first stage of the hiring process and does not measure actual hiring, the provision of information about a service does not measure actual service provision, although it is still the first step. Finally, when researchers send out limited templates, as discussed in the introduction, they may be subject to template bias—that is, their findings will only be externally valid for the groups that the templates describe.

Ethics

Ethical concerns with audit studies derive from academic concerns about deception to negative spillovers on constituents by taking away time from and potentially providing incorrect information to public servants. While it is beyond the scope of this paper to provide a full framework of the ethics of audit studies, a general guiding principle is that the benefits to audit studies should always be greater than any potential harm. In practice, this principle generally means that researchers need to be careful of participant time, make sure data are anonymized so that no participant can be called out, and if there is a chance of potential harm from providing false information then the response should be followed up with a post-consent form. If these precautions cannot be taken, then an audit study may not be an appropriate tool. New methods of combining audit technologies with real, rather than hypothetical, stimuli are a promising path forward.

While deception is common in psychology experiments, it is frowned upon in economics experiments if the subjects know they are participating in a study. Experimental economists believe that their subject pool should never try to second guess the purpose of the experiment (e.g., Davis & Holt 1993, Ortman & Hertwig 2002). Deception in any economics experiment in which subjects know they are participating would have negative spillovers on this goal. Audit studies circumvent this problem because the participants generally do not know that they are participating in the experiment (McClendon 2012; James et al. 2017).

There is some evidence that the subject pool of politicians and constituents do not find these types of experiments to be problematic (Naurin & Öhberg 2021). That said, deception is commonly considered unethical and is considered by some to have unknown societal costs including to political behaviors (Findley & Nielson 2016; James et al. 2017). Findley & Nielson (2016) reports a pair of studies showing that some degree of deception is necessary for unbiased results that may provide considerable policy implications. They also provide suggestions on how to reduce (but not necessarily eliminate) deception including informed consent and recruiting confederates that write the stimuli using their own names and contact information to avoid identity deception. Landgrave (2020) echoes the suggestion for using confederates, and proposes a novel study design that avoids identity, activity, and misinformation deception by stating up front that the contact is for research and then analyzing responses interacted with the incentivization characteristics.

Ethically, the researcher has an obligation to minimize the harm done to unwitting participants. Experiments that require a lengthy response should only be done if the benefit outweighs the cost. Data should be anonymized—no individual participant should be able to be called out as biased (this warning is particularly important for public university researchers who may be subject to FOIA requests for their research data). With PA experiments, in addition to the direct costs to the participants, there may be negative spillovers if responding to the audit study distracts participants from real constituent requests and duties or causes them to inaccurately change their beliefs about constituent priorities. These concerns should be included in the cost-benefit calculation.

Decisions to post-consent participants should be weighed carefully—on the one hand, post-consenting makes participants aware of the deception of the experiment and could make participants suspicious of future interactions with real constituents. It is possible that an agency discovering the deception could additionally provide negative reputational consequences to the research team and university.⁶ On the other hand, participants gather information from and can be otherwise affected by the deceptive communications. If it is likely they will make decisions based on these communications or suffer any harm from believing the deception, then the post-consent is necessary to mitigate that harm. The post-consent should also include the ability for participants to remove their data, giving participants a locus-of-control if they did not want to participate.

Perhaps the ideal audit study reduces deception and/or loss of participant time by combining an experiment with real outputs that will be used for another purpose (e.g., Block et al. 2021; Grose et al. 2015; Landgrave 2020; Peisakhin & Pinto 2010). A recent study from economics pairs real college students requesting advice from working professionals to study the effects of gender on career advice (Gallen & Wasserman 2021). In PA, this type of combination could translate to FOIA audit studies requesting information that they need for another study (e.g., Jenkins et al. 2020; Spáč et al. 2018). Similarly, researchers studying the effect of the personal demographics of fundraisers could partner with a non-profit that benefits from the donations (Jilke et al. 2019; Landgrave 2020). In addition, working with existing organizations is one way researchers planning audit studies can find collaborators and focus research on topics relevant to political agents. For example, while not an audit study because the

⁶ Gaddis et al. (2021) discuss the cost-benefit analysis of ethical concerns in more detail.

focus was on changing behavior rather than measuring it, Bergner, Desmarais, & Hird (2019) increased the impact of their field experiment by partnering with an advocacy group.

V. The future of audits

As email and electronic messaging platforms become the primary modes of communication, both the ease of performing an audit study and their potential scope continue to grow. Even though some of the increased ease of communication may be of limited use in PA (e.g., Google ads would be an ineffective medium to study elite lawmakers), there will be an overall increase in feasibly explorable theories and questions about practices in PA. There will be new opportunities to reach different offices of governing agencies, broader coverage of governing bodies with smaller geographic reach (e.g., county or city governments, such as all Spáč et al. 2018 which studied all Slovak municipalities), and, on the other side of the scale, increased ease of international studies (e.g., 11 countries are compared in Magni & Ponce de Leon 2021). Reaching sufficient sample size for expected effect sizes (e.g., 10,268 government representatives were emailed in Crawfurd and Ramli 2022) and engaging a sufficient variety of participants (e.g., both elected officials and the general public were studied in Block et al. 2021) should become more manageable with increased use of digital communication. By reaching more participants, the time cost for individual participants can be minimized, for example, sending fewer stimuli per participant or dividing one complex “ask” into multiple stages with different participants per stage.

Digital communications also make sending multiple stimuli to the same participant simple, inexpensive, and quick, so within-subjects and follow-up studies are also getting easier to perform with broader scope. Within-subjects studies can provide more statistical power that may be necessary when the pool of participants is small (e.g., political elites), while follow-up studies (e.g., Butler & Crabtree 2017) investigate the effects of an initial stimulus on the participant’s response to a subsequent stimulus.

Simultaneously, the ongoing trend towards digital communications favors more intricate stimuli designed to detect how responses might be influenced by more complex relationships between stimulus and participant characteristics. Various audits have shown that intersections between characteristics in a single stimulus (e.g., race crossed with age) or between the stimulus and the participant (e.g., gender of stimulus crossed with gender of participant) are important (e.g., Habel & Birch 2019; Druckman & Shafranek 2020; Rhinehart 2020). Studies with sufficient participants to support diversification of the stimuli will benefit from the increased ease of creating and distributing balanced stimuli that explore more such interactions, thus providing a fuller exploration of the market.

In other fields, audit studies have commonly been used to explore the hiring market and the housing market, typically exploring discrimination based on some characteristic that varies between stimuli, such as using the name to indicate race. These fields have recently called for broader use of the audit study method to understand different markets. At the same time, these fields have demonstrated that researchers need to be more careful that their characteristics actually signal what they think is signaled. Additionally, there has been a push to go beyond documenting existence of discrimination to

identifying its mechanisms, e.g., through side-by-side use of an audit study and a lab experiment or survey exploring hypotheticals, or by taking advantage of natural experiments (Gaddis 2018). Each of these three areas are of note for the future use of audit studies in PA: explore new markets, be careful about internal and external validity, and push for a deeper understanding of the market.

Works Cited

- Adman, P., & Jansson, H. (2017). A field experiment on ethnic discrimination among local Swedish public officials . *Local Government Studies*, 43(1), 44-63.
<https://doi.org/10.1080/03003930.2016.1244052>
- Alizade, J., Dancygier, R., & Ditzmann, R. K. (2021). National Penalties Reversed: The Local Politics of Citizenship and Politician Responsiveness to Immigrants. *Journal of Politics*, 83(3), 867-883. <https://doi.org/10.1086/711056>
- Barlow, M. R., & Lahey, J. N. (2018). What Race Is Lacey? Intersecting Perceptions of Racial Minority Status and Social Class. *Social Science Quarterly*, 99(5), 1680-1698.
- Ben-Aaron, J., Denny, M., Desmarais, B., & Wallach, H. (2017). Transparency by Conformity: A Field Experiment Evaluating Openness in Local Government. *Public Administration Review*, 77(1), 68-77.
- Bergner, C., Desmarais, B. A. & Hird, J. 2019. Speaking truth in power: Scientific evidence as motivation for policy activism. *Journal of Behavioral Public Administration*, 2(1), 1-11.
- Bertrand, M., & Duflo, E. (2017). Field experiments on discrimination. *Handbook of Economic Field Experiments*, 1, 309-393.
- Block, R., Crabtree, C., Holbein, J. B., & Monson, J. Q. (2021). Are Americans less likely to reply to emails from Black people relative to White people? *Proceedings of the National Academy of Sciences*, 118(52), e2110347118. <https://doi.org/10.1073/pnas.2110347118>
- Bol, D., Gschwend, T., Zittel, T., & Zittlau, S. (2021). The importance of personal vote intentions for the responsiveness of legislators: A field experiment. *European Journal of Political Research*, 60(2), 455-473. <https://doi.org/10.1111/1475-6765.12408>
- Bouwman, R., & Grimmelikhuijsen, S. (2016). Experimental public administration from 1992 to 2014. *The International Journal of Public Sector Management*, 29(2), 110-131.
<http://dx.doi.org/10.1108/IJPSM-07-2015-0129>
- Broockman, D. 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. <https://doi.org/10.1111/ajps.12018>
- Broockman, D. E., & Green, D. P. (2014). Do Online Advertisements Increase Political Candidates' Name Recognition or Favorability? Evidence from Randomized Field Experiments. *Political Behavior*, 36(2), 263-289.
- Butler, D. M., & Broockman, D. E. (2011). Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators. *American Journal of Political Science*, 55(3), 463-477.
- Butler, D. M., & Crabtree, C. (2017). Moving Beyond Measurement: Adapting Audit Studies to Test Bias-Reducing Interventions. *Journal of Experimental Political Science*, 4(1), 57-67. <https://doi.org/10.1017/XPS.2017.11>
- Butler, D. M., De Vries, C. E., & Solaz, H. (2019). Studying policy diffusion at the individual level: Experiments on nationalistic biases in information seeking. *Research & Politics*, 6(4), 2053168019891619. <https://doi.org/10.1177/2053168019891619>

- Butler, D. M., Karpowitz, C. F., & Pope, J. C. (2012). A Field Experiment on Legislators' Home Styles: Service versus Policy. *The Journal of Politics*, 74(2), 474-486.
<https://doi.org/10.1017/s0022381611001708>
- Button, P., Dils, E., Harrell, B., Fumarco, L., & Schwegman, D. (2020). *Gender Identity, Race, and Ethnicity Discrimination in Access to Mental Health Care: Preliminary Evidence from a Multi-Wave Audit Field Experiment* (No. w28164). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28164/w28164.pdf
- Carnes, N., & Holbein, J. (2019). Do public officials exhibit social class biases when they handle casework? Evidence from multiple correspondence experiments. *PLOS ONE*, 14(3), e0214244. <https://doi.org/10.1371/journal.pone.0214244>
- Chen, J., Pan, J., & Xu, Y. (2016). Sources of Authoritarian Responsiveness: A Field Experiment in China. *American Journal of Political Science*, 60(2), 383-400.
<https://doi.org/10.1111/ajps.12207>
- Coppock, A. (2019). Avoiding Post-Treatment Bias in Audit Experiments. *Journal of Experimental Political Science*, 6(1), 1-4. <https://doi.org/10.1017/XPS.2018.9>
- Costa, M. (2017). How Responsive are Political Elites? A Meta-Analysis of Experiments on Public Officials. *Journal of Experimental Political Science (JEPS)*, 4(3), 241-254.
- Crabtree, C., Kim, J. Y., Gaddis, S. M., Holbein, J. B., Guage, C., & Marx, W. W. (2023). Validated names for experimental studies on race and ethnicity. *Scientific Data*, 10(1), 130. <https://doi.org/10.1038/s41597-023-01947-0>
- Crawfurd, L., & Ramlí, U. (2022). Discrimination by politicians against religious minorities: Experimental evidence from the UK. *Party Politics*, 28(5), 826-833.
<https://doi.org/10.1177/13540688211021053>
- Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2015). Do Employers Prefer Workers Who Attend For-Profit Colleges? Evidence from a Field Experiment. *Journal of Policy Analysis & Management*, 34(4), 881-903. <https://doi.org/10.1002/pam.21863>
- Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton University Press.
- De Vries, C. , Dinas, E., & Solaz, H. (2016). You Have Got Mail! How Intrinsic and Extrinsic Motivations Shape Constituency Service in the European Parliament. *Working Paper*. https://irihs.ihs.ac.at/id/eprint/3955/1/wp_140.pdf
- Dhima, K. (2022). Do Elites Discriminate against Female Political Aspirants? Evidence from a Field Experiment. *Politics & Gender*, 18(1), 126-157.
<https://doi.org/10.1017/S1743923X20000227>
- Distelhorst, G., & Hou, Y. (2014). Ingroup Bias in Official Behavior: A National Field Experiment in China. *Quarterly Journal of Political Science*, 9(2), 203-230.
<https://doi.org/10.1561/100.00013110>
- Driscoll, A., Cepaluni, G., Guimaraes, F. d. S., & Spada, P. (2018). Prejudice, Strategic Discrimination, and the Electoral Connection: Evidence from a Pair of Field Experiments in Brazil. *American Journal of Political Science*, 62(4), 781.
<https://doi.org/10.1111/ajps.12371>
- Dropp, K., & Peskowitz, Z. (2012). Electoral Security and the Provision of Constituency Service. *The Journal of Politics*, 74(1), 220-234. <https://doi.org/10.1017/s0022381611001216>

- Dynes, A. M., Hassell, H. J. G., & Miles, M. R. (2021). Political Ambition and Constituent Service: Does Ambition Influence How Local Officials Respond to Electoral and Non-Electoral Service Requests? *American Politics Research*, 49(6), 646-654.
<https://doi.org/10.1177/1532673X211013479>
- Einstein, K. L., & Glick, D. M. (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. <https://doi.org/10.1111/ajps.12252>
- Epstein, B., Bode, L., & Connolly, J. M. (2021). Do squeaky wheels get the grease? Understanding when and how municipalities respond to online requests. *New Media & Society*, 14614448211031249. <https://doi.org/10.1177/14614448211031249>
- Ernst, R., Nguyen, L., & Taylor, K. C. (2013). Citizen Control: Race at the Welfare Office. *Social Science Quarterly*, 94(5), 1283-1307. <https://doi.org/10.1111/ssqu.12013>
- Fix, M., & Struyk, R. J. (1993). *Clear and convincing evidence : measurement of discrimination in America*. Urban Institute Press.
- Gaddis, S. M. (2017a). How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies. *Sociological Science*, 4, 469-489.
<https://doi.org/10.15195/v4.a19>
- Gaddis, S. M. (2017b). Racial/Ethnic Perceptions from Hispanic Names: Selecting Names to Test for Discrimination. *Socius*, 3, 2378023117737193.
<https://doi.org/10.1177/2378023117737193>
- Gaddis, S. M. (2018). An Introduction to Audit Studies in the Social Sciences. In S. M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 3-44). Springer International Publishing. https://doi.org/10.1007/978-3-319-71153-9_1
- Gaddis, S. M., Larsen, E., Crabtree, C., & Holbein, J. (2021). Discrimination against Black and Hispanic Americans is highest in hiring and housing contexts: A meta-analysis of correspondence audits. *PNAS*. <http://www.pnas.org/cgi/doi/10.1073/pnas.XXXXXXXXXXX>
- Gaddis, S. M., Crabtree, C., ..., & Weisshaar, K. (2021, March 25). Sending politicians fake emails is sometimes necessary. *Times Higher Education*.
<https://www.timeshighereducation.com/blog/sending-politicians-fake-emails-sometimes-necessary>
- Gaikwad, N., & Nellis, G. (2021). Do Politicians Discriminate Against Internal Migrants? Evidence from Nationwide Field Experiments in India. *American Journal of Political Science*, 65(4), 790-806. <https://doi.org/10.1111/ajps.12548>
- Gallen, Y., Wasserman, M. (2021). *Informed Choices: Gender Gaps in Career Advice* (14072). IZA Institute of Labor Economics.
- Gell-Redman, M., Visalvanich, N., Crabtree, C., & Fariss, C. J. (2018). It's All about Race: How State Legislators Respond to Immigrant Constituents. *Political Research Quarterly*, 71(3), 517-531.
<https://doi.org/10.1177/1065912917749322>
- Giger, N., Lanz, S., & de Vries, C. (2020). The motivational basis of constituency work: how intrinsic and extrinsic motivations interact . *Political Science Research & Methods*, 8(3), 493-508. <https://doi.org/10.1017/psrm.2019.19>

- Giner-Sorolla, R. (2018, January 24). Powering Your Interaction. *Approaching Significance: A methodology blog for social psychology*.
<https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/>
- Golder, S. N., Crabtree, C., & Dhima, K. (2019). Legislative representation and gender (bias). *Political Science*, 71(1), 1-16. <https://doi.org/10.1080/00323187.2019.1632151>
- Grohs, S., Adam, C., & Knill, C. (2016). Are Some Citizens More Equal than Others? Evidence from a Field Experiment. *Public Administration Review*, 76(1), 155-164.
- Grose, C. R., Malhotra, N., & Van Houweling, R. P. (2015). Explaining Explanations: How Legislators Explain their Policy Positions and How Citizens React. *American Journal of Political Science*, 59(3), 724-743. <http://www.jstor.org/stable/24583093>
- Habel, P., & Birch, S. (2019). A Field Experiment on the Effects of Ethnicity and Socioeconomic Status on the Quality of Representation. *Legislative Studies Quarterly*, 44(3), 389-420. <https://doi.org/10.1111/lsq.12230>
- Hansen, J. A., & Tummers, L. (2020). A Systematic Review of Field Experiments in Public Administration. *Public Administration Review*, 80(6), 921-931. <https://doi.org/10.1111/puar.13181>
- Hayes, T. J., & Bishin, B. G. (2020). Do elected officials serve the poor on health care? Evidence from a field experiment on members of congress and state legislators. *The Social Science Journal*, 1-17. <https://doi.org/10.1080/03623319.2020.1783963>
- Hemker, J., & Rink, A. (2017). Multiple Dimensions of Bureaucratic Discrimination: Evidence from German Welfare Offices. *American Journal of Political Science*, 61(4), 786-803. <https://doi.org/10.1111/ajps.12312>
- Hou, Y., Liu, C., & Crabtree, C. (2020). Anti-muslim bias in the Chinese labor market. *Journal of Comparative Economics*, 48(2), 235-250. <https://doi.org/10.1016/j.jce.2019.12.001>
- Hughes, D. A., Gell-Redman, M., Crabtree, C., Krishnaswami, N., Rodenberger, D., & Monge, G. (2020). Persistent Bias Among Local Election Officials. *Journal of Experimental Political Science*, 7(3), 179-187. <https://doi.org/10.1017/XPS.2019.23>
- James, O., John, P., & Moseley, A. (2017). *Experiments in Public Management Research: Challenges and Contributions*. Cambridge University Press. <https://doi.org/10.1017/9781316676912.006>
- Janusz, A., & Lajevardi, N. (2016). The Political Marginalization of Latinos: Evidence from Three Field Experiments. Working Paper.
- Jenkins, N. R., Landgrave, M., & Martinez, G. E. (2020). Do political donors have greater access to government officials? Evidence from a FOIA field experiment with US municipalities. *Journal of Behavioral Public Administration*, 3(2). <https://doi.org/10.30636/jbpa.32.111>
- Jilke, S., Lu, J., Xu, C., & Shinohara, S. (2019). Using Large-Scale Social Media Experiments in Public Administration: Assessing Charitable Consequences of Government Funding of Nonprofits. *Journal of Public Administration Research & Theory*, 29(4), 627-639. <https://doi.org/10.1093/iopart/muy021>
- John, P., & Blume, T. (2018). How best to nudge taxpayers? The impact of message simplification and descriptive social norms on payment rates in a central London local authority. *Journal of Behavioral Public Administration*, 1. <https://doi.org/10.30636/jbpa.11.10>

- Kalla, J., Rosenbluth, F., & Teele, D. L. (2018). Are You My Mentor? A Field Experiment on Gender, Ethnicity, and Political Self-Starters. *Journal of Politics*, 80(1), 336-341. <https://doi.org/10.1086/693984>
- Kalla, J. L., & Broockman, D. E. (2016). Campaign Contributions Facilitate Access to Congressional Officials: A Randomized Field Experiment. *American Journal of Political Science*, 60(3), 545-558. <https://doi.org/10.1111/ajps.12180>
- Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened Résumés. *Administrative Science Quarterly*, 61(3), 469-502. <https://doi.org/10.1177/0001839216639577>
- Lagunes, P., & Pocasangre, O. (2019). Dynamic transparency: An audit of Mexico's Freedom of Information Act. *Public Administration*, 97(1), 162-176. <https://doi.org/10.1111/padm.12553>
- Lahey, J., & Beasley, R. A. (2009). Computerizing audit studies. *Journal of Economic Behavior & Organization*, 70(3), 508-514. <https://EconPapers.repec.org/RePEc:eee:jeborg:v:70:y:2009:i:3:p:508-514>
- Lahey, J., & Beasley, R. (2018). Technical Aspects of Correspondence Studies. In S. M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 81-101). Springer International Publishing. https://doi.org/10.1007/978-3-319-71153-9_4
- Lajevardi, N. (2020). Access denied: exploring Muslim American representation and exclusion by state legislators. *Politics, Groups, and Identities*, 8(5), 957-985. <https://doi.org/10.1080/21565503.2018.1528161>
- Landgrave, M. (2020). Can We Reduce Deception in Elite Field Experiments? Evidence from a Field Experiment with State Legislative Offices. *State Politics & Policy Quarterly*, 20(4), 489-507. <https://doi.org/10.1177/1532440020925723>
- Landgrave, M. (2021). Do Politicians Ethnically Discriminate Against Hispanics? Evidence from a Field Experiment with State Legislative Offices. *Legislative Studies Quarterly*, 46(3), 621. <https://doi.org/10.1111/lsq.12311>
- Landgrave, M., & Weller, N. (2022). Do Name-Based Treatments Violate Information Equivalence? Evidence from a Correspondence Audit Experiment. *Political Analysis*, 30(1), 142-148. <https://doi.org/10.1017/pan.2020.52>
- Li, H., & Van Ryzin, G. G. (2017). A Systematic Review of Experimental Studies in Public Management Journals. In G. G. Van Ryzin, O. James, & S. R. Jilke (Eds.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 20-36). Cambridge University Press. <https://doi.org/DOI: 10.1017/9781316676912.003>
- Linos, E. (2018). More Than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police. *Journal of Public Administration Research & Theory*, 28(1), 67-85. <https://doi.org/10.1093/jopart/mux032>
- Lippens, L., Vermeiren, S., & Baert, S. (2023). The State of Hiring Discrimination: A Meta-Analysis of (Almost) All Recent Correspondence Experiments. *European Economic Review*, 151, 104315. <https://doi.org/10.1016/j.eurocorev.2022.104315>
- Loewen, P. J., & MacKenzie, M. K. (2019). Service Representation in a Federal System: A Field Experiment. *Journal of Experimental Political Science*, 6(2), 93-107. <https://doi.org/10.1017/XPS.2018.21>
- Lowande, K., & Proctor, A. (2020). Bureaucratic Responsiveness to LGBT Americans. *American Journal of Political Science*, 64(3), 664-681. <https://doi.org/10.1111/ajps.12493>

- Lowande, K., Ritchie, M., & Lauterbach, E. (2019). Descriptive and Substantive Representation in Congress: Evidence from 80,000 Congressional Inquiries. *American Journal of Political Science*, 63(3), 644-659. <https://doi.org/10.1111/ajps.12443>
- Mackenzie-Liu, M., Schwegman, D. J., & Lopoo, L. M. (2020). Do Foster Care Agencies Discriminate Against Gay Couples? Evidence from a Correspondence Study. *Journal of Policy Analysis and Management*, 40(4), 1140-1170. <https://doi.org/10.1002/pam.22237>
- Magni, G., & Ponce de Leon, Z. (2021). Women Want an Answer! Field Experiments on Elected Officials and Gender Bias. *Journal of Experimental Political Science*, 8(3), 273-284. <https://doi.org/10.1017/XPS.2020.23>
- Margetts, H. (2011). Experiments for Public Management Research. *Public Management Review*, 13(2), 189-208.
- McClendon, G. H. (2012). Ethics of Using Public Officials As Field Experiment Subjects. *Newsletter of the APSA Experimental Section*, 3(1), 13-20. <https://sites.google.com/site/gwynethmcclendon/>
- McClendon, G. H. (2016). Race and Responsiveness: An Experiment with South African Politicians. *Journal of Experimental Political Science*, 3(1), 60-74. <https://doi.org/10.1017/XPS.2015.10>
- Mendez, M. S. (2014). Who Represents the Interests of Undocumented Latinos? A Study of State Legislators. *Working Paper*.
- Mendez, M. S., & Grose, C. R. (2018). Doubling Down: Inequality in Responsiveness and the Policy Preferences of Elected Officials. *Legislative Studies Quarterly*, 43(3), 457-491. <https://doi.org/10.1111/lsq.12204>
- Michelson, M. R. (2006). Mobilizing the Latino Youth Vote: Some Experimental Results. *Social Science Quarterly* 87(5): 1188-1206.
- Michener, G., Velasco, R. B., Contreras, E., & Rodrigues, K. F. (2020). Googling the requester: Identity-questing and discrimination in public service provision. *Governance*, 33(2), 249-267. <https://doi.org/10.1111/gove.12416>
- Mikula, Š., & Montag, J. (2022). Roma and Bureaucrats: A Field Experiment in the Czech Republic. *Muni Econ Working Paper Series* 2022(01).
- Naurin, E., & Öhberg, P. (2021). Ethics in Elite Experiments: A Perspective of Officials and Voters. *British Journal of Political Science*, 51(2), 890-898.
- Neiman, J. (2017). The impact of education on legislative responsiveness in three field experiments. *Cogent Social Sciences*, 3(1), 1282034. <https://doi.org/10.1080/23311886.2017.1282034>
- Neumark, D. (2018). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature*, 56(3), 799-866. <https://doi.org/10.1257/jel.20161309>
- Newland, S. A., & Liu, J. C.-E. (2021). Ethnic identity and local government responsiveness in Taiwan. *Governance*, 34(3), 875-892. <https://doi.org/10.1111/gove.12546>
- Nyhan, B., & Reifler, J. (2015). The Effect of Fact-Checking on Elites: A Field Experiment on U.S. State Legislators. *American Journal of Political Science*, 59(3), 628-640. <http://www.jstor.org/stable/24583087>
- Olsen, A. L., Kyhse-Andersen, J. H., & Moynihan, D. (2022). The Unequal Distribution of Opportunity: A National Audit Study of Bureaucratic Discrimination in Primary School

- Access. *American Journal of Political Science*, 66(3), 587-603.
<https://doi.org/10.1111/ajps.12584>
- Ortmann, A., & Hertwig, R. (2002). The Costs of Deception: Evidence from Psychology. *Experimental Economics*, 5(2), 111-131.
- Peisakhin, L. (2012). Transparency and Corruption: Evidence from India. *Journal of Law and Economics*, 55(1), 129-149. [https://doi.org/0022-2186/2012/5501-0005\\$10.00](https://doi.org/0022-2186/2012/5501-0005$10.00)
- Peisakhin, L., & Pinto, P. (2010). Is transparency an effective anti-corruption strategy? Evidence from a field experiment in India. *Regulation & Governance*, 4(3), 261-280.
<https://doi.org/10.1111/j.1748-5991.2010.01081.x>
- Pfaff, S., Crabtree, C., Kern, H. L., & Holbein, J. B. (2021). Do Street-Level Bureaucrats Discriminate Based on Religion? A Large-Scale Correspondence Experiment among American Public School Principals. *Public Administration Review*, 81(2), 244-259.
<https://doi.org/10.1111/puar.13235>
- Phillips, D. (2019). Do Comparisons of Fictional Applicants Measure Discrimination When Search Externalities are Present? Evidence from Existing Experiments. *The Economic Journal*, 129(621), 2240-2264.
<https://EconPapers.repec.org/RePEc:oup:econjl:v:129:y:2019:i:621:p:2240-2264>.
- Porter, E., & Rogowski, J. C. (2018). Partisanship, Bureaucratic Responsiveness, and Election Administration: Evidence from a Field Experiment. *Journal of Public Administration Research & Theory*, 28(4), 602-617. <https://doi.org/10.1093/jopart/muy025>
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, 114(41), 10870-10875.
- Rhinehart, S. (2020). Mentoring the Next Generation of Women Candidates: A Field Experiment of State Legislators. *American Politics Research*, 48(4), 492-505.
<https://doi.org/10.1177/1532673X20906459>
- Rodríguez, R. P., & Rossel, C. (2018). A field experiment on bureaucratic discretionary bias under FOI laws. *Government Information Quarterly*, 35(3), 418-427.
<https://doi.org/10.1016/j.giq.2018.06.001>
- Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3), 523-534. <https://doi.org/10.1016/j.labeco.2009.04.005>
- Schultz, W., Khazian, A., & Zaleski, A. (2008). Using normative social influence to promote conservation among hotel guests. *Social Influence*, 3(1), 4-23.
<https://doi.org/10.1080/15534510701755614>
- Spáč, P., Voda, Petr, Zagrapan, Jozef. (2018). Does the freedom of information law increase transparency at the local level? Evidence from a field experiment. *Government Information Quarterly*, 35, 408-417. <https://doi.org/10.1016/j.giq.2018.05.003>.
- Taghizadeh, J. L. (2021). Caseload, time-pressure and discrimination: Are school principals who have less time for e-mail correspondence more likely to discriminate in their e-mail replies? *Journal of Behavioral Public Administration*, 4(2).
<https://doi.org/10.30636/jbpa.42.271>
- Taghizadeh, J. L., Åström, A., & Adman, P. (2022). Do Politicians Discriminate against Ethnic Minority Constituents? : A Field Experiment on Social Interactions between Citizens and

- Swedish Local Politicians. *Parliamentary Affairs*, 75(1), 154-172.
<https://doi.org/10.1093/pa/gsaa053>
- Terechshenko, Z., Crabtree, C., Eck, K., & Fariss, C. J. (2019). Evaluating the influence of international norms and shaming on state respect for rights: an audit experiment with foreign embassies. *International Interactions*, 45(4), 720-735.
<https://doi.org/10.1080/03050629.2019.1622543>
- Thomsen, D. M., & Sanders, B. K. (2020). Gender Differences in Legislator Responsiveness. *Perspectives on Politics*, 18(4), 1017-1030. <https://doi.org/10.1017/S1537592719003414>
- Timm, J. (2021). Wielding Constituent Power? The effect of electoral threats on responsiveness and response quality in grassroots policy lobbying. Working Paper.
- Vaishnav, M., Khosla, S., Milliff, A., & Osnos, R. (2019). Digital India? an email experiment with Indian legislators. *India Review*, 18(3), 243-263.
<https://doi.org/10.1080/14736489.2019.1616259>
- Vuolo, M., Lageson, S., & Uggen, C. (2017). Criminal Record Questions in the Era of 'Ban the Box'. *Criminology & Public Policy*, 16(1), 139-165. <https://doi.org/10.1111/1745-9133.12250>
- White, A. R., Nathan, N. L., & Faller, J. K. (2015). What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials. *American Political Science Review*, 109(1), 129-142. <https://doi.org/10.1017/S0003055414000562>
- Worthy, B., John, P., & Vannoni, M. (2017). Transparency at the Parish Pump: A Field Experiment to Measure the Effectiveness of Freedom of Information Requests in England. *Journal of Public Administration Research & Theory*, 27(3), 485-500.
<https://doi.org/10.1093/jopart/muw063>

Appendix

Systematic Literature Review Papers

- Adman, P., & Jansson, H. (2017). A field experiment on ethnic discrimination among local Swedish public officials. *Local Government Studies*, 43(1), 44-63.
<https://doi.org/10.1080/03003930.2016.1244052>
- Alizade, J., Dancygier, R., & Ditlmann, R. K. (2020). National Penalties Reversed: The Local Politics of Citizenship and Politician Responsiveness to Immigrants. *The Journal of Politics*, 83(3), 867-883. <https://doi.org/10.1086/711056>
- Bárcena Juárez, S. A., & Kerevel, Y. P. (2021). Communicative responsiveness in the Mexican Senate: a field experiment. *Democratization*, 28(8), 1387-1405.
<https://doi.org/10.1080/13510347.2021.1907348>
- ben-Aaron, J., Denny, M., Desmarais, B., & Wallach, H. (2017). Transparency by Conformity: A Field Experiment Evaluating Openness in Local Government. *Public Administration Review*, 77(1), 68-77.
- Block, R., Crabtree, C., Holbein John, B., & Monson, J. Q. (2021). Are Americans less likely to reply to emails from Black people relative to White people? *Proceedings of the National Academy of Sciences*, 118(52), e2110347118. <https://doi.org/10.1073/pnas.2110347118>
- Bol, D., Gschwend, T., Zittel, T., & Zittlau, S. (2021). The importance of personal vote intentions for the responsiveness of legislators: A field experiment. *European Journal of Political Research*, 60(2), 455-473. <https://doi.org/10.1111/1475-6765.12408>
- Broockman, D. 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. <https://doi.org/10.1111/ajps.12018>
- Brown, J. R., & Hilbig, H. (2021). Locked Out of College: When Admissions Bureaucrats Do and Do Not Discriminate. *British Journal of Political Science*, 1-11.
<https://doi.org/10.1017/S0007123420000848>
- Butler, D. M., & Broockman, D. E. (2011). Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators. *American Journal of Political Science*, 55(3), 463-477.
- Butler, D. M., & Crabtree, C. (2017). Moving Beyond Measurement: Adapting Audit Studies to Test Bias-Reducing Interventions. *Journal of Experimental Political Science*, 4(1), 57-67. <https://doi.org/10.1017/XPS.2017.11>
- Butler, D. M., De Vries, C. E., & Solaz, H. (2019). Studying policy diffusion at the individual level: Experiments on nationalistic biases in information seeking. *Research & Politics*, 6(4), 2053168019891619. <https://doi.org/10.1177/2053168019891619>
- Butler, D. M., Karpowitz, C. F., & Pope, J. C. (2012). A Field Experiment on Legislators' Home Styles: Service versus Policy. *The Journal of Politics*, 74(2), 474-486.
<https://doi.org/10.1017/s0022381611001708>

- Carnes, N., & Holbein, J. (2019). Do public officials exhibit social class biases when they handle casework? Evidence from multiple correspondence experiments. *PLOS ONE*, 14(3), e0214244. <https://doi.org/10.1371/journal.pone.0214244>
- Chen, J., Pan, J., & Xu, Y. (2016). Sources of Authoritarian Responsiveness: A Field Experiment in China. *American Journal of Political Science*, 60(2), 383-400. <https://doi.org/10.1111/ajps.12207>
- Crawfurd, L., & Ramli, U. (2022). Discrimination by politicians against religious minorities: Experimental evidence from the UK. *Party Politics*, 28(5), 826-833. <https://doi.org/10.1177/13540688211021053>
- Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2015). Do Employers Prefer Workers Who Attend For-Profit Colleges? Evidence from a Field Experiment. *Journal of Policy Analysis & Management*, 34(4), 881-903. <https://doi.org/10.1002/pam.21863>
- Dhima, K. (2022). Do Elites Discriminate against Female Political Aspirants? Evidence from a Field Experiment. *Politics & Gender*, 18(1), 126-157. <https://doi.org/10.1017/S1743923X20000227>
- Dinesen, P. T., Dahl, M., & Schiøler, M. (2021). When Are Legislators Responsive to Ethnic Minorities? Testing the Role of Electoral Incentives and Candidate Selection for Mitigating Ethnocentric Responsiveness. *American Political Science Review*, 115(2), 450-466. <https://doi.org/10.1017/S0003055420001070>
- Distelhorst, G., & Hou, Y. (2014). Ingroup Bias in Official Behavior: A National Field Experiment in China. *Quarterly Journal of Political Science*, 9(2), 203-230. <https://doi.org/10.1561/100.00013110>
- Distelhorst, G., & Hou, Y. (2017). Constituency Service under Non-democratic Rule: Evidence from China. *The Journal of Politics*, 79(3), 1024-1040. <https://doi.org/10.1086/690948>
- Ditlmann, R. K., & Lagunes, P. (2014). The (Identification) Cards You Are Dealt: Biased Treatment of Anglos and Latinos Using Municipal-Issued versus Unofficial ID Cards. *Political Psychology*, 35(4), 539-555.
- Driscoll, A., Cepaluni, G., Guimaraes, F. d. S., & Spada, P. (2018). Prejudice, Strategic Discrimination, and the Electoral Connection: Evidence from a Pair of Field Experiments in Brazil. *American Journal of Political Science*, 62(4), 781. <https://doi.org/10.1111/ajps.12371>
- Dropp, K., & Peskowitz, Z. (2012). Electoral Security and the Provision of Constituency Service. *The Journal of Politics*, 74(1), 220-234. <https://doi.org/10.1017/s0022381611001216>
- Druckman, J. N., & Shafranek, R. M. (2020). The Intersection of Racial and Partisan Discrimination: Evidence from a Correspondence Study of Four-Year Colleges. *Journal of Politics*, 82(4), 1602-1606. <https://doi.org/10.1086/708776>
- Dynes, A. M., Hassell, H. J. G., & Miles, M. R. (2021). Political Ambition and Constituent Service: Does Ambition Influence How Local Officials Respond to Electoral and Non-Electoral Service Requests? *American Politics Research*, 49(6), 646-654. <https://doi.org/10.1177/1532673X211013479>
- Einstein, K. L., & Glick, D. M. (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. <https://doi.org/10.1111/ajps.12252>

- Epstein, B., Bode, L., & Connolly, J. M. (2021). Do squeaky wheels get the grease? Understanding when and how municipalities respond to online requests. *New Media & Society*, 14614448211031249. <https://doi.org/10.1177/14614448211031249>
- Ernst, R., Nguyen, L., & Taylor, K. C. (2013). Citizen Control: Race at the Welfare Office. *Social Science Quarterly*, 94(5), 1283-1307. <https://doi.org/10.1111/ssqu.12013>
- Fang, A. H., Guess, A. M., & Humphreys, M. (2019). Can the Government Deter Discrimination? Evidence from a Randomized Intervention in New York City. *Journal of Politics*, 81(1), 127-141. <https://doi.org/10.1086/700107>
- Fenton, J., & Stephens-Dougan, L. (2021). Are Black state legislators more responsive to emails associated with the NAACP versus BLM? A field experiment on Black intragroup politics. *The Journal of Race, Ethnicity, and Politics*, 1-16. <https://doi.org/10.1017/rep.2021.13>
- Findley, M. G., Nielson, D. L., & Sharman, J. C. (2015). Causes of Noncompliance with International Law: A Field Experiment on Anonymous Incorporation. *American Journal of Political Science*, 59(1), 146-161. <https://doi.org/10.1111/ajps.12141>
- Gaikwad, N., & Nellis, G. (2021). Do Politicians Discriminate Against Internal Migrants? Evidence from Nationwide Field Experiments in India. *American Journal of Political Science*, 65(4), 790-806. <https://doi.org/10.1111/ajps.12548>
- Gell-Redman, M. a. V., Neil and Crabtree, Charles and Fariss, Christopher J. (2018). It's all about race: how State legislators respond to immigrant constituents. *Political Research Quarterly*, 71(3), 517-531 <https://doi.org/10.1177/1065912917749322>
- Giger, N., Lanz, S., & de Vries, C. (2020). The motivational basis of constituency work: how intrinsic and extrinsic motivations interact. *Political Science Research & Methods*, 8(3), 493-508. <https://doi.org/10.1017/psrm.2019.19>
- Giulietti, C., Tonin, M., & Vlassopoulos, M. (2019). Racial Discrimination in Local Public Services: A Field Experiment in the United States. *Journal of the European Economic Association*, 17(1), 165-204. <https://doi.org/10.1093/jeea/jvx045>
- Golder, S. N., Crabtree, C., & Dhima, K. (2019). Legislative representation and gender (bias). *Political Science*, 71(1), 1-16. <https://doi.org/10.1080/00323187.2019.1632151>
- Grimmelikhuijsen, S., John, P., Meijer, A., & Worthy, B. (2019). Do freedom of information laws increase transparency of government? A replication of a field experiment. *Journal of Behavioral Public Administration*, 2. <https://doi.org/10.30636/jbpa.12.34>
- Grohs, S., Adam, C., & Knill, C. (2016). Are Some Citizens More Equal than Others? Evidence from a Field Experiment. *Public Administration Review*, 76(1), 155-164.
- Grose, C. R., Malhotra, N., & Parks Van Houweling, R. (2015). Explaining Explanations: How Legislators Explain their Policy Positions and How Citizens React. *American Journal of Political Science*, 59(3), 724-743. <https://doi.org/10.1111/ajps.12164>
- Habel, P., & Birch, S. (2019). A Field Experiment on the Effects of Ethnicity and Socioeconomic Status on the Quality of Representation. *Legislative Studies Quarterly*, 44(3), 389-420. <https://doi.org/10.1111/lsq.12230>
- Hayes, T. J., & Bishin, B. G. (2020). Do elected officials serve the poor on health care? Evidence from a field experiment on members of congress and state legislators. *The Social Science Journal*, 1-17. <https://doi.org/10.1080/03623319.2020.1783963>

- Hemker, J., & Rink, A. (2017). Multiple Dimensions of Bureaucratic Discrimination: Evidence from German Welfare Offices. *American Journal of Political Science*, 61(4), 786-803. <https://doi.org/10.1111/ajps.12312>
- Hou, Y., Liu, C., & Crabtree, C. (2020). Anti-muslim bias in the Chinese labor market. *Journal of Comparative Economics*, 48(2), 235-250. <https://doi.org/10.1016/j.jce.2019.12.001>
- Hughes, D. A., Gell-Redman, M., Crabtree, C., Krishnaswami, N., Rodenberger, D., & Monge, G. (2020). Persistent Bias Among Local Election Officials. *Journal of Experimental Political Science*, 7(3), 179-187. <https://doi.org/10.1017/XPS.2019.23>
- Jenkins, N. R., Landgrave, M., & Martinez, G. E. (2020). Do political donors have greater access to government officials? Evidence from a FOIA field experiment with US municipalities. *Journal of Behavioral Public Administration*, 3(2). <https://doi.org/10.30636/jbpa.32.111>
- Jensen, N. M., Findley, M. G., & Nielson, D. L. (2020). Electoral Institutions and Electoral Cycles in Investment Incentives: A Field Experiment on Over 3,000 U.S. Municipalities. *American Journal of Political Science*, 64(4), 807-822. <https://doi.org/10.1111/ajps.12499>
- Jilke, S., Lu, J., Xu, C., & Shinohara, S. (2019). Using Large-Scale Social Media Experiments in Public Administration: Assessing Charitable Consequences of Government Funding of Nonprofits. *Journal of Public Administration Research & Theory*, 29(4), 627-639. <https://doi.org/10.1093/jopart/muy021>
- Jilke, S., Van Dooren, W., & Rys, S. (2018). Discrimination and Administrative Burden in Public Service Markets: Does a Public-Private Difference Exist? . *Journal of Public Administration Research & Theory*, 28(3), 423-439. <https://doi.org/10.1093/jopart/muy009>
- Kalla, J., Rosenbluth, F., & Teele, D. L. (2018). Are You My Mentor? A Field Experiment on Gender, Ethnicity, and Political Self-Starters. *Journal of Politics*, 80(1), 336-341. <https://doi.org/10.1086/693984>
- Kalla, J. L., & Broockman, D. E. (2016). Campaign Contributions Facilitate Access to Congressional Officials: A Randomized Field Experiment. *American Journal of Political Science*, 60(3), 545-558. <https://doi.org/10.1111/ajps.12180>
- Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened Résumés. *Administrative Science Quarterly*, 61(3), 469-502. <https://doi.org/10.1177/0001839216639577>
- Lagunes, P., & Pocasangre, O. (2019). Dynamic transparency: An audit of Mexico's Freedom of Information Act. *Public Administration*, 97(1), 162-176. <https://doi.org/10.1111/padm.12553>
- Lajevardi, N. (2020). Access denied: exploring Muslim American representation and exclusion by state legislators. *Politics, Groups, and Identities*, 8(5), 957-985. <https://doi.org/10.1080/21565503.2018.1528161>
- Landgrave, M. (2020). Can We Reduce Deception in Elite Field Experiments? Evidence from a Field Experiment with State Legislative Offices. *State Politics & Policy Quarterly*, 20(4), 489-507. <https://doi.org/10.1177/1532440020925723>
- Landgrave, M. (2021). Do Politicians Ethnically Discriminate Against Hispanics? Evidence from a Field Experiment with State Legislative Offices. *Legislative Studies Quarterly*, 46(3), 621. <https://doi.org/10.1111/lsg.12311>

- Landgrave, M., & Weller, N. (2022). Do Name-Based Treatments Violate Information Equivalence? Evidence from a Correspondence Audit Experiment. *Political Analysis*, 30(1), 142-148. <https://doi.org/10.1017/pan.2020.52>
- Lloren, A. (2017). Does direct democracy increase communicative responsiveness? A field experiment with Swiss politicians. *Research & Politics*, 4(1), 2053168017700738. <https://doi.org/10.1177/2053168017700738>
- Loewen, P. J., & MacKenzie, M. K. (2019). Service Representation in a Federal System: A Field Experiment. *Journal of Experimental Political Science*, 6(2), 93-107. <https://doi.org/10.1017/XPS.2018.21>
- Lowande, K., & Proctor, A. (2020). Bureaucratic Responsiveness to LGBT Americans. *American Journal of Political Science*, 64(3), 664-681. <https://doi.org/10.1111/ajps.12493>
- Magni, G., & Ponce de Leon, Z. (2021). Women Want an Answer! Field Experiments on Elected Officials and Gender Bias. *Journal of Experimental Political Science*, 8(3), 273-284. <https://doi.org/10.1017/XPS.2020.23>
- McClendon, G. H. (2016). Race and Responsiveness: An Experiment with South African Politicians. *Journal of Experimental Political Science*, 3(1), 60-74. <https://doi.org/10.1017/XPS.2015.10>
- Mendez, M. S., & Grose, C. R. (2018). Doubling Down: Inequality in Responsiveness and the Policy Preferences of Elected Officials. *Legislative Studies Quarterly*, 43(3), 457-491. <https://doi.org/10.1111/lsq.12204>
- Michelson, M. R. (2006). Mobilizing the Latino Youth Vote: Some Experimental Results. *Social Science Quarterly*, 87(5), 1188-1206. <http://www.jstor.org/stable/42956602>
- Michener, G., Velasco, R. B., Contreras, E., & Rodrigues, K. F. (2020). Googling the requester: Identity-questing and discrimination in public service provision. *Governance*, 33(2), 249-267. <https://doi.org/10.1111/gove.12416>
- Moy, B. J. (2021). Can Social Pressure Foster Responsiveness? An Open Records Field Experiment with Mayoral Offices. *Journal of Experimental Political Science*, 8(2), 117-127. <https://doi.org/10.1017/XPS.2020.22>
- Neiman, J. (2017). The impact of education on legislative responsiveness in three field experiments. *Cogent Social Sciences*, 3(1), 1282034. <https://doi.org/10.1080/23311886.2017.1282034>
- Newland, S. A., & Liu, J. C.-E. (2021). Ethnic identity and local government responsiveness in Taiwan. *Governance*, 34(3), 875-892. <https://doi.org/10.1111/gove.12546>
- Oberfield, Z. W., & Incantalupo, M. B. (2021). Racial Discrimination and Street-Level Managers: Performance, Publicness, and Group Bias. *Public Administration Review*, 81(6), 1055-1070. <https://doi.org/10.1111/puar.13376>
- Olsen, A. L., Kyhse-Andersen, J. H., & Moynihan, D. (2022). The Unequal Distribution of Opportunity: A National Audit Study of Bureaucratic Discrimination in Primary School Access. *American Journal of Political Science*, 66(3), 587-603. <https://doi.org/10.1111/ajps.12584>
- Peisakhin, L. (2012). Transparency and Corruption: Evidence from India. *The Journal of Law and Economics*, 55(1), 129-149. <https://doi.org/10.1086/663727>

- Peisakhin, L., & Pinto, P. (2010). Is transparency an effective anti-corruption strategy? Evidence from a field experiment in India. *Regulation & Governance*, 4(3), 261-280.
<https://doi.org/10.1111/j.1748-5991.2010.01081.x>
- Pfaff, S., Crabtree, C., Kern, H. L., & Holbein, J. B. (2021). Do Street-Level Bureaucrats Discriminate Based on Religion? A Large-Scale Correspondence Experiment among American Public School Principals. *Public Administration Review*, 81(2), 244-259.
<https://doi.org/10.1111/puar.13235>
- Poole, E. G. (2019). How Institutional Culture Trumps Tier Effects: Evidence from Government Responsiveness to FOI Requests. *Journal of Public Administration Research & Theory*, 29(2), 210-226. <https://doi.org/10.1093/jopart/muy039>
- Porter, E., & Rogowski, J. C. (2018). Partisanship, Bureaucratic Responsiveness, and Election Administration: Evidence from a Field Experiment. *Journal of Public Administration Research & Theory*, 28(4), 602-617. <https://doi.org/10.1093/jopart/muy025>
- Rhinehart, S. (2020). Mentoring the Next Generation of Women Candidates: A Field Experiment of State Legislators. *American Politics Research*, 48(4), 492-505.
<https://doi.org/10.1177/1532673X20906459>
- Rochlitz, M., Mitrokhina, E., & Nizovkina, I. (2021). Bureaucratic discrimination in electoral authoritarian regimes: Experimental evidence from Russia. *European Journal of Political Economy*, 66, 101957. <https://doi.org/10.1016/j.ejpoleco.2020.101957>
- Rodríguez, R. P., & Rossel, C. (2018). A field experiment on bureaucratic discretionary bias under FOI laws. *Government Information Quarterly*, 35(3), 418-427.
<https://doi.org/10.1016/j.giq.2018.06.001>
- Southern, R., & Purdam, K. (2016). The changing representation interface: democracy and direct contact with politicians. *Journal of Civil Society*, 12(1), 101-120.
<https://doi.org/10.1080/17448689.2016.1134388>
- Spáč, P., Voda, Petr, Zagrapan, Jozef. (2018). Does the freedom of information law increase transparency at the local level? Evidence from a field experiment. *Government Information Quarterly*, 35, 408-417. <https://doi.org/10.1016/j.giq.2018.05.003>.
- Sun, T. (2019). Deliberate Differentiation by the Chinese State: Outsourcing Responsibility for Governance. *The China Quarterly*, 240, 880-905.
<https://doi.org/10.1017/S0305741019000420>
- Taghizadeh, J. L. (2021). Caseload, time-pressure and discrimination: Are school principals who have less time for e-mail correspondence more likely to discriminate in their e-mail replies? *Journal of Behavioral Public Administration*, 4(2).
<https://doi.org/10.30636/jbpa.42.271>
- Taghizadeh, J. L., Åström, A., & Adman, P. (2022). Do Politicians Discriminate against Ethnic Minority Constituents? A Field Experiment on Social Interactions between Citizens and Swedish Local Politicians. *Parliamentary Affairs*, 75(1), 154-172.
<https://doi.org/10.1093/pa/gsaa053>
- Terechshenko, Z., Crabtree, C., Eck, K., & Fariss, C. J. (2019). Evaluating the influence of international norms and shaming on state respect for rights: an audit experiment with foreign embassies. *International Interactions*, 45(4), 720-735.
<https://doi.org/10.1080/03050629.2019.1622543>

- Thomsen, D. M., & Sanders, B. K. (2020). Gender Differences in Legislator Responsiveness. *Perspectives on Politics*, 18(4), 1017-1030. <https://doi.org/10.1017/S1537592719003414>
- Vaishnav, M., Khosla, S., Milliff, A., & Osnos, R. (2019). Digital India? an email experiment with Indian legislators. *India Review*, 18(3), 243-263. <https://doi.org/10.1080/14736489.2019.1616259>
- Vuolo, M., Lageson, S., & Uggen, C. (2017). Criminal Record Questions in the Era of 'Ban the Box'. *Criminology & Public Policy*, 16(1), 139-165. <https://doi.org/10.1111/1745-9133.12250>
- Wagner, A. J. (2021). Piercing the veil: Examining demographic and political variables in state FOI Law administration. *Government Information Quarterly*, 38(1), 101541. <https://doi.org/10.1016/j.giq.2020.101541>
- White, A. R., Nathan, N. L., & Faller, J. K. (2015). What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials. *American Political Science Review*, 109(1), 129-142. <https://doi.org/10.1017/S0003055414000562>
- Wiener, E. (2020). Getting a High Heel in the Door: An Experiment on State Legislator Responsiveness to Women's Issue Lobbying. *Political Research Quarterly*, 74(3), 729-743. <https://doi.org/10.1177/1065912920939186>
- Worthy, B., John, P., & Vannoni, M. (2017). Transparency at the Parish Pump: A Field Experiment to Measure the Effectiveness of Freedom of Information Requests in England . *Journal of Public Administration Research & Theory*, 27(3), 485-500. <https://doi.org/10.1093/jopart/muw063>