AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Bureaucratic Responsiveness to LGBT Americans

Kenneth Lowande University of Michigan
Andrew Proctor University of Minnesota

Abstract: *Marriage rights were extended to same-sex couples in the United States in 2015. However, anecdotes of bureaucratic noncompliance (in the form of bias or denial of license issuance) raise the possibility that de jure marriage equality has not led to equality in practice. We investigate this by conducting a nationwide audit experiment of local-level marriage license–granting officials in the United States. These officials vary in the constituencies they serve, as well as how they are selected, allowing us to evaluate long-standing hypotheses about bureaucratic responsiveness. Overall, we find no evidence of systematic discrimination against same-sex couples—regardless of responsiveness measure, institutions, ideology, or prior state legal history. We find, however, that among same-sex couples, officials tended to be more responsive to lesbian couples. In contrast to evidence in other areas of service provision, such as policing and federal assistance programs, we find bureaucrats tasked with provision of marriage services show little evidence of discrimination.*

Verification Materials: The data and materials required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/JYKL9M.

After the nationwide legalization of same-sex marriage in the United States in 2015, Kim Davis, a Democratic county clerk in Kentucky, refused to provide a license to a same-sex couple. This paralleled other anecdotes of discrimination, including that of then State Supreme Court Justice Roy Moore in Alabama, who ordered county officials to cease issuing marriage licenses to same-sex couples altogether.[1] These cases raise a perennial question in democratic governance: whether the officials who must implement law will comply with directives from judges and legislators—who have limited coercive means of enforcement.

These concerns are particularly acute after watershed moments in policy change, in which bureaucrats have substantial discretion by virtue of their distance from formal principals. Executive departments failed to implement Title VI of the Civil Rights Act of 1964 for years after passage (Minta 2011). There is evidence that uneven enforcement of the Voting Rights Act of 1965 continues to influence the incorporation of underrepresented minorities in elections (Marschall and Rutherford 2016). Most notably, the implementation of the Supreme Court's *Brown v. Board of Education* (347 U.S. 483, 1954) decision required decades-long action on the part of prosecutors, activists, legislators, and even presidents (Daugherity and Bolton 2008).

Similarly, although *Obergefell v. Hodges* (576 U.S. __, 2015) changed the status of same-sex marriages overnight in 14 states, the decision had to be implemented by over 6,000 public officials with vastly different constituencies,

Kenneth Lowande is Assistant Professor, Department of Political Science, University of Michigan, 5700 Haven Hall, 505 South State Street, Ann Arbor, MI 48109 (lowande@umich.edu). Andrew Proctor is Presidential Post-Doctoral Associate, Department of Political Science, University of Minnesota, 267 19th Ave S., Minneapolis, MN 55455 (proct061@umn.edu).

Previous versions were presented at the 2019 annual meetings of the Southern Political Science Association, Austin, TX, and the Midwest Political Science Association, Chicago, IL. We thank Andrew Clarke, Charles Crabtree, Thomas Gray, Hans Hassell, Adam Hughes, George Krause, Adam Seth Levine, Scott Limbocker, Noah Nathan, Jennifer Selin, Dara Strolovitch, Omar Wasow, and participants at the Princeton Research in Experimental Social Science (PRESS) workshop for helpful comments and suggestions. Special thanks to Justin Fortney and Sebastian Leder Macek for research assistance. PRESS and the Center for the Study of Democratic Politics at Princeton University provided funding. This study was approved by institutional review boards at Princeton University (IRB# 10522) and the University of Michigan (IRB# 154110).

[1]"See Alabama Chief Justice Orders Halt to Same-Sex Marriage Licenses," *Reuters*, January 6, 2016, https://www.reuters.com/article/us-alabama-gaymarriage-idUSKBN0UK2AR20160106.

American Journal of Political Science, Vol. 64, No. 3, July 2020, Pp. 664–681

 DOI: 10.1111/ajps.12493

political principals, and little to no formal oversight. To investigate the responsiveness of bureaucrats to LGBT Americans, we conduct a national audit experiment of marriage license–granting officials in the United States. Specifically, we investigate whether street-level bureaucrats are less responsive to same-sex couples when they request information about the process of obtaining a marriage license.

Measuring differential service provision across sexual orientation has important consequences for scholarship on LGBT politics, executive accountability, and public policy. Our systematic analysis of bureaucratic responsiveness to LGBT people in the United States takes place after nationwide legalization. Thus, beyond the descriptive task of measuring potential discrimination, our study investigates whether bureaucrats comply with major policy change. In addition, since the rules governing the selection of license-granting officials vary by locality, our analysis speaks to long-standing questions about whether the accountability mechanisms of public officials influence service provision. Institutional differences in responsiveness are usually explored in the context of legislative and judicial decision makers. However, most citizen–government interactions involve executive officials, so it is important to understand how these mechanisms impact service provision. Finally, since states vary in their prior legality of same-sex marriage and other anti-LGBT laws, we assess the impact of the historical legacy of these laws on local compliance. The historical influence of geographic context on local racial attitudes (e.g., Acharya, Blackwell, and Sen 2016), for example, suggests these laws may have long-term impacts on the treatment of same-sex couples.

Overall, in contrast to persistent evidence that public officials are less responsive to marginalized groups (e.g., Gell-Redman et al. 2018; Mendez and Grose 2018), we find little evidence of systematic differences across sexual orientations of putative couples. Specifically, the 95% confidence intervals of differences in response rates, three measures of response quality, and even congratulatory language are not distinguishable from zero, are inconsistently signed, and typically do not include a magnitude effect greater than 3 percentage points. These estimates are relatively stable across estimation procedures and consistent across selection institutions and state legal history. We find some evidence that officials were less responsive to gay men relative to lesbian couples; however, this effect is largely driven by the fact that officials were systemically less responsive to male email senders, regardless of partner. Our results are consistent with emerging evidence in some areas of public policy that public officials concerned primarily with service provision exhibit less evidence of bias when dealing with putative citizens (e.g., Einstein and Glick 2017; Jilke, Van Dooren, and Rys 2018; Porter and Rogowski 2018). More broadly, our evidence does not suggest that localized anecdotes of discrimination are systematically representative.

Same-Sex Marriage and Bureaucratic Responsiveness

Same-sex marriage in the United States is an ideal case for addressing proposed mechanisms for bureaucratic responsiveness. Since a large majority of citizens are married at least once, it poses a broadly applicable administrative barrier. Classic agency models predict behavioral differences based on the selection mechanism of officials. Scholars have long argued that courts play an important role in influencing bureaucratic behavior (e.g., Wood and Waterman 1994). Though bureaucratic compliance with judicial directives is generally high, there is some evidence it is conditioned by agents' policy preferences (Spriggs 1996). Research on other historically marginalized groups suggests systematic differences in service provision. The historical moment raises questions about long-term bureaucratic compliance with judicial directives. Importantly, these differences are difficult to identify in an observational setting. A dearth of centralized data sources and underreporting of discrimination cases necessitates an audit experiment.

This approach is common, but past work does not address the questions we raise. Much work investigates whether public officials are more or less responsive to marginalized groups, typically racial and ethnic minority constituents. In studies requesting information about voting, scholars find that black and Latino constituents receive fewer replies and less accurate information from public officials than putatively white constituents (Butler and Broockman 2011; White, Nathan, and Faller 2015). This is consistent with other studies that find blacks and Latinos face discrimination in the labor market (e.g., Bertrand and Mullainathan 2003; Pager, Bonikowski, and Western 2009), as consumers (Ayres and Siegelman 1995; Doleac and Stein 2013; Turner et al. 2013), and in higher education (Milkman, Akinola, and Chugh 2014). A notable exception to these findings is Einstein and Glick (2017), who find mixed evidence for racial discrimination among public housing officials.

These studies are informative about the nature and degree of discrimination facing racial and ethnic minorities, but they cannot speak to whether other marginalized constituencies, such as lesbian, gay, bisexual, and transgender (LGBT) people, face similar forms of discrimination when interacting with public officials. The first known audit experiment examining discrimination based on sexual orientation was conducted by Barry Adam in 1981 (Badgett et al. 2008). More recent work has argued sexual minorities face discrimination in the labor market (Bailey, Wallace, and Wright 2013; Crow et al. 1991; Hebl et al. 2002; Weichselbaumer 2003). Van Hoye and Lievens (2003) provide an exception. Other audit experiments find same-sex couples receive differential treatment compared to opposite-sex couples in public accommodations (Jones 1996; Walters and Curran 1996). Although these studies are suggestive, they tend to be severely underpowered, show small effects that are inconsistent across outcome measures, or involve treatment conditions that strain credulity.[2]

More importantly, this existing research focuses on private discrimination and was conducted prior to important changes in American politics. Most audit studies of sexual orientation discrimination were published in or before 2003, around the time that same-sex marriage became the central organizing issue of the LGBT rights movement in the United States. Since then, studies have documented changes in public attitudes about LGBT people and their rights (Pew Research Center 2013, 2017). Moreover, there have been significant changes surrounding the legality of same-sex marriage in the United States. Prior to 2004, most states did not have bans on same-sex marriage written into law, although same-sex couples did not have access to the institution of marriage. After Massachusetts became the first state to legalize same-sex marriage, a social conservative countermovement led to an increase in the number of states with legal bans from 3 to 30 between 2004 and 2012. Over the same period, the number of states with legal same-sex marriage increased from 1 to 8.

In statewide referendums in 2012, Washington, Maryland, and Maine voted to legalize same-sex marriage and Minnesota rejected a ban on same-sex marriage. One year later, the Supreme Court ruled part of the Defense of Marriage Act unconstitutional in *United States v. Windsor* (570 U.S. 744, 2013), granting legally married same-sex couples equal status under federal law. Between 2013 and 2015, the number of states with bans on same-sex marriage decreased to 14. These final 14 states were eventually forced to legalize same-sex marriage when the Court ruled all bans on same-sex marriage unconstitutional in *Obergefell v. Hodges.*[3]

Nonetheless, de jure extension of marriage rights in the United States did not end political controversy surrounding the rights of same-sex couples. Since the *Obergefell* decision, opponents reframed the debate about LGBT rights as violations of the freedom of religion. The most famous example is Kim Davis, a Democratic county clerk in Kentucky who refused to provide a marriage license to a same-sex couple in 2015 on such grounds (Blinder and Perez-Pena 2015). The state of Kentucky then amended its law to exempt clerks from signing marriage licenses (SB021620.1000 - 1695 - 8234). Similarly, Roy Moore of the Alabama State Supreme Court directed counties to defy the U.S. Supreme Court ruling (Robertson 2016). As of 2018, there were eight counties in Alabama that no longer grant marriage licenses, to avoid issuing licenses to same-sex couples (Dunigan 2018). In Texas and New York, same-sex couples have reported being denied marriage licenses by county clerks as recently as July 2018, more than 3 years after the legalization of same-sex marriage nationwide (Prager 2018; Sanchez 2018). In West Virginia, a couple reported a marriage license–granting official had called them an "abomination" (Wooston and Somashekhar 2017). This suggests public officials may be systematically less responsive to same-sex couples. Though this is our primary expectation, we also consider two mechanisms for heterogeneous effects.

Historical Legacy

First, the weight of evidence in the social sciences suggests that past institutions may have persistent impacts on present service provision. A large and active literature examines the long-term impacts of legal history on economic development (for a review, see Nunn 2009). In American politics, the lasting effects of (defunct) political institutions are felt across a variety of contexts. For example, Acharya, Blackwell, and Sen (2016) argue slavery has lasting impacts on political attitudes, specifically racial animus and conservatism. Another recent

[2]In Hebl et al. (2002), for example, volunteers were sent to apply for jobs wearing hats that read "gay and proud." In other studies, fictitious résumés include work histories that may signal both sexual orientation and variation in experience.

[3]Although states are the actors that define and set marriage laws, the Defense of Marriage Act was enacted at the federal level in 1998. The law passed with veto-proof majorities and was signed by President Bill Clinton after the Hawaii State Supreme Court ruled that bans on same-sex marriage constituted gender discrimination in 1997. Section 2 of the law defined the federal government's position on the legality of same-sex marriage by only affording federal marriage rights and benefits to heterosexual couples. Section 3 of the law allowed states to not recognize legal same-sex marriages that were performed in other states (or countries). Section 2 was ruled unconstitutional in *Windsor*, and Section 3 was ruled unconstitutional in *Obergefell.*

example is Trounstine (2018), who demonstrates prior housing institutions impact economic inequality and reinforce contemporary segregation.

There are numerous reasons suggesting laws that discriminate against same-sex couples will have similar effects. Anecdotally, with the exception of New York, all reported cases of marriage license discrimination against LGBT couples occurred in states with overturned same-sex marriage bans or anti-sodomy laws. Moreover, marriage license granting shares institutional features that are thought to aid in historical transmission. Housing policies, for example, also saw sweeping nondiscriminatory changes to policy ordered via judicial directive—but these were implemented by a highly decentralized system of private actors and local governments. Finally, the political context that created them—including the attitudes of the population from which officials are drawn—cannot be expected to dissipate overnight.

Selection Rules and Responsiveness

Existing audit studies of public officials typically examine contexts with little or no institutional variation. In the case of marriage, however, licenses are granted locally by a wide variety of institutions, including county clerks, municipal clerks, probate judges, clerks of wills, county registrars, luxury resorts, and centralized state offices. In total, there are more than 6,000 marriage license–issuing authorities in the United States.

For our purposes, one key factor to consider is that some marriage license–issuing authorities are appointed, whereas others are elected. Accountability mechanisms like these are thought to shape the behavior and responsiveness of public officials (Kimball and Kropf 2006). Research has found conditional evidence that local election officials are more responsive to copartisans in some contexts, although means of selection did not matter to responsiveness (Porter and Rogowski 2018). Other work finds that selection institution shapes the behavior of judges (Canes-Wrone, Clark, and Kelly 2014). When judges are elected, they are responsive to voter ideology (Lim 2013) and alter their sentencing behavior in response to electoral incentives (Huber and Gordon 2004). When their reappointment is conditional on approval from politicians, judges alter their behavior to align with the preferences of the legislators (Gray 2017) and governors (Gray 2019). Similarly, studies of regulatory commissions have found that selection method influences the behavior of bureaucrats (Besley and Coate 2003; Fields, Klein, and Sfiridis 1997).

This suggests marriage license–issuing authorities may vary in responsiveness depending on their method of selection. Specifically, we expect that elected bureaucrats will be less responsive to same-sex couples in conservative jurisdictions, as these authorities align their behavior with the preferences of their constituents. In contrast, we expect that appointed bureaucrats will respond to same-sex and opposite-sex couples at similar rates since they do not face electoral incentives to respond to constituency public opinion.

Summary of Expectations

In summary, the sudden change in policy, discretion of officials, anecdotes of discrimination, and long-term public attitude change suggest that same-sex marriage is an ideal case for studying the "street-level" implications of major judicial policy changes. We list our expectations below.

> *Discrimination Hypothesis:* Licensers will be less responsive to same-sex requesters.
>
> *Legacy Hypothesis:* Licensers in states with a recent legal history of limiting LGBT rights will be less responsive to same-sex requesters, relative to officials in other states.
>
> *Selection Hypothesis:* Licensers who are elected will be less responsive to same-sex requesters in areas of higher local-level conservatism, compared with other requesters, relative to elected officials in liberal areas and appointed officials.

The legacy and selection hypotheses imply two- and three-way interaction effects, respectively. We discuss power more extensively when contextualizing the results. However, it should be noted that our study will only speak to whether there is evidence of substantively large effects for these more demanding hypotheses.

Research Design

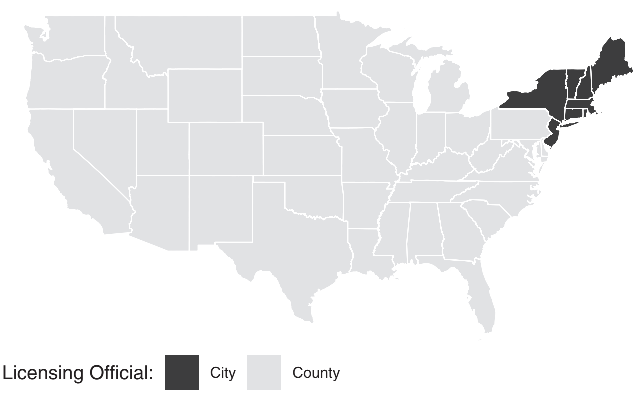
We conducted a randomized experiment with a straightforward 2×2 design. The marriage license–granting official received an inquiry from a $Requester = \{Female, Male\}$, inquiring on behalf of his or her future $Spouse = \{Female, Male\}$. Table 1 reports this setup. Accordingly, our discrimination hypothesis will be evaluated by comparing responsiveness to G_1 and G_4 with that of G_2 and G_3. Other expectations will be tested with difference-in-difference estimates among

TABLE 1 Treatment Groups

		Requester	
		Female	***Male***
Spouse	*Female*	G_1 (1,103)	G_2 (1,105)
	Male	G_3 (1,105)	G_4 (1,103)

Note: *N* of 4,414 is less jurisdictions missing electronic contact info.

FIGURE 1 Marriage License-Granting Institutions in the United States



sample subgroups. We have no ex ante expectations about differences across gay and lesbian couples, but our design is capable of measuring them by comparing G_1 and G_4.

We expect heterogeneous treatment effects and have reason to believe some types of officials will be more responsive than others. Therefore, we employ multivariate, continuous blocking to improve balance (Moore 2012; Moore and Schnakenberg 2016). Specifically, we block based on five covariates: service level (county or city; see Figure 1), institution type (elected or appointed), contact means (email or online form), local ideology, and population.[4] Local ideology is measured with the average Democratic share of the two-party presidential vote from 2008 and 2012. Population estimates were obtained from the Census Bureau. Subnational variation in the level of marriage-licensing authorities and the sparsity of population in some counties necessitate blocking on population. Moreover, local population is correlated with ideology because of the influence of urban city centers. Thus, we were concerned about uncovering a spurious relationship between ideology and responsiveness. Urban authorities might be more responsive, for example, because they are more professionalized and accustomed to handling such inquiries.

Sample

Our subject population is 5,123 jurisdictions that issue marriage licenses in the 48 contiguous states. In Hawaii, marriage licenses are granted by numerous private individuals and businesses (hotels and resorts, in particular) authorized by the state. We have not acquired the full registry of these grantors, who number in the thousands. Moreover, since they are not public officials, they are outside the scope of our study. In Alaska, licenses are granted by cities, counties, and the state, and the application process is centralized. The District of Columbia, likewise, operates a single marriage bureau. We exclude all of these outliers. We must also exclude officials who do not have electronic contact information posted online, which leaves 4,414 cases. Contact information came in the form of an email or an online contact form. Notably, we believe we overestimate the true number of jurisdictions, as our list is obtained by assuming every city in applicable states (even those with fewer than 50 residents and no webpage) issues marriage licenses. Many of the missing cases may be outside the scope conditions of our study. However, missingness of electronic contact information was found to be positively correlated with local ideology. We defer discussion of the influence of these excluded jurisdictions on our analysis to the discussion and supporting information (SI; see SI Appendix A).

Though the title of the officials who issue licenses varies by state (clerk, recorder, registrar, judge, etc.), the most basic distinction between them is the level of government that issues the license. Figure 1 plots this regional variation—as states in the Northeast allow cities to issue licenses.

Instrument

The experimental instrument is an email inquiry from a fictional constituent (Figure 2). To increase the likelihood that the official received the treatment, the email was short and the treatment appeared in the first nine words. This follows other studies that find treatment effects for named, putative constituents (e.g., Butler and Broockman 2011; White, Nathan, and Faller 2015). We also avoided declarative statements about identity and randomly vary

[4]We suspected that officials with online contact forms, as well as those operating at the city level, would be slower to respond. There is little research on online contact forms, but officials unwilling to post direct contact information suggested they were less likely to be responsive. Costa (2017) found evidence that national-level officials were more responsive than subnational officials. We extended this logic to the county–city comparison.

FIGURE 2 Email Instrument Example

```
Hello,

My name is [random: {female full name, male full name}]. My future
[random: {husband, wife}], [random: {female first name, male first name}],
and I need to get a marriage license. When can we apply for one?
How long is it valid? How much does it cost?

We plan to tie the knot on [date].

Thank you,
[rand: {female first name, male first name}]
```

salutation and email structure to reduce the likelihood of subject discovery (Butler and Crabtree 2017). This addresses a potential concern in prior research showing discrimination against same-sex couples. First names were chosen to provide a clear signal of gender. These were Brandon, Dylan, Jacob, Elizabeth, Jessica, and Megan. To increase our confidence that the official received the treatment, we also included a gendered partner synonym. In addition, we chose surnames that have consistent interpretation as white across regions, to prevent variation in attitudes about race and ethnicity from influencing our findings (Crabtree and Chykina 2018). These were Anderson, Nielson, and Walsh. This also means that the results should be regarded as limited in scope to whites. Although we believe there may be important interactions between race and sexual orientation, we leave this for future research for two reasons. First, we have no theoretical priors about heterogeneity extension of LGBT marriage rights by race and ethnicity. Second, power considerations prevented us from fully crossing sexual orientation with race and gender in the design.

One advantage of the policy chosen for study is that the rules determining cost, expiration, valid applicants, and other features vary by state and, in some cases, locality. Thus, inquiries like these are common, and responses about these details provide a measure of quality. This is important, as the quality of the information provided is arguably as important as receiving a response. In addition, many webpages did not contain this information. In fact, some webpages indicate that marriage licenses are granted via appointment and thus *require* correspondence to schedule. Other officials indicated (via email) that the information provided on their webpage was inaccurate or outdated. In short, the realism of the instrument presents an additional advantage that avoids the pitfalls of asking for information already duplicated online.

One potential concern, however, is that discrimination in this context may require additional effort or be obscured by highly routinized bureaucratic behavior. In addition, past work suggests differential responsiveness in service provision is more common in face-to-face interactions (e.g., Maynard-Moody and Musheno 2003; Pepinsky, Pierskalla, and Sacks 2017). Although it is important to note that some forms of discrimination in email correspondence require officials to exert effort, we have reason to believe that this case is not particularly prone to concealing differences in service provision. First, the simplest form of differential treatment would be to not reply, or to forward the constituent to a webpage. So discrimination can require *less* effort. In addition, though some emails were undoubtedly drafted from templates, a large share of replies contains informal or personal details, apologizes for response delays, asks follow-up questions, or only addresses the specific questions asked in the inquiry. This suggests that differences would not be obscured by boilerplate responses.

Second, our study is comparable to numerous other audit studies that find differential treatment in routine inquiries. Butler and Broockman (2011) and White, Nathan, and Faller (2015) both find differential responsiveness to short inquiries asking local officials about voting requirements. Distelhorst and Hou (2014) and Einstein and Glick (2017) find differences in responses to inquiries asking for information about requirements for government benefits. Each of these studies uncovers evidence of discrimination against racial/ethnic minorities and asks for information in a routine service context. Notably, discrimination that is not simply nonresponse sometimes required additional effort in these studies.

Implementation

We sent email messages to all validated addresses and contact forms in late summer 2018. We selected this time because it was at least a month in advance of the fictitious wedding in September, which is among the most popular months for a wedding in the United States.[5] The emails were sent from six accounts over 2 days to prevent being marked as spam. The contact forms were manually entered by researchers those same days. The data collection process had a terminal date on a Saturday in late September 2018, which was the latest wedding date provided in the outgoing emails. Any responses after this date were coded as nonresponsive.

No response suggested discovery of the experiment. In fact, in some cases, the lack of a follow-up response from the fictitious emailer by the week of the wedding date was alarming enough to provoke additional follow-up emails by officials. There were nine jurisdictions that had to be dropped for stable unit treatment value assumption (SUTVA) violations because different treatments were sent to the same email address by mistake. There is no consistent pattern in these cases. The outgoing emails were sent at different times. One jurisdiction responded to both emails differently. Two jurisdictions responded to the first, but not the second. Two jurisdictions responded to both identically. Others did not reply to either. Importantly, although these mistakes are unfortunate, they are insufficiently numerous to pose a problem.

Externalized Costs and Ethical Considerations

In addition to standard ethical considerations raised in past work, we provide estimates of the externalized cost of the study. It is important to note that the experiment was reviewed and approved by two institutional review boards. Still, past research discusses the ethics of deception and lack of informed consent particular to audit experiments of public officials (e.g., Butler and Broockman 2011; Einstein and Glick 2017). Consistent with this research, and because of the routine and informational nature of the inquiry, we argue that our study minimizes the potential harm and burden placed on participants. However, we also argue that providing a more precise estimate of the total burden placed on participants is appropriate. The prominence of the research we cite and the reduced cost of fielding have led to the proliferation of audit experiments of public officials in the social sciences. Potential long-term consequences are not difficult to imagine: increased skepticism of constituent communication by public officials, or aggregate costs that exceed the benefit of new knowledge claims.

For the purposes of this study, we define externalized costs as the total public expense siphoned by the experiment—meaning the costs are external *to the researcher*. The most obvious way to operationalize these costs is to estimate the time our study took away from other work. In this case, we received 3,285 responses. The responses, including all salutations and email signatures, contained 460,433 words, or about 140 words per reply. Estimates of average typing speed in the United States suggest each of these replies took about 3 or 3.4 minutes on average. If we use the 2017 median hourly wage for license clerks according to the Census, this means a conservative (i.e., ceiling) estimate for the total external cost of the study was $3,355.92, or about 187 hours of work from public officials, nationally. This is likely conservative since some replies included text that was part of a template response. With the same simple arithmetic, the cost of *reading* would be $266.89—assuming all sent emails that did not bounce were read. Of course, these are not invoices. By providing them, we argue that researchers conducting similar experiments should be precise about externalized costs, and we illustrate that minor interventions impose some "burden" in the aggregate.

In this case, we argue that the externalized cost of this study is trivial when compared to the social benefits of learning about this phenomenon. First, there is a dearth of social scientific research on discrimination against LGBT people, even as cases of reported discrimination continue. Second, our research design improves on what few studies on LGBT discrimination exist. Third, the unique historical moment provides an opportunity to learn about political processes of general interest—namely, bureaucratic compliance with judicial directives. Finally, since our selection and legacy hypotheses imply covariate interactions, we could not limit the sample of jurisdictions selected for the audit. The reduction in power precluded the possibility of attempting to reduce the number of officials contacted. Thus, in our view, we have minimized the costs associated with this study while maximizing the study's social value—and the latter greatly outweighs the former.

Response Coding

We received 3,285 responses from 2,971 unique jurisdictions. We are interested in variation in the responsiveness of public officials to the fictional inquiry. We measure this several ways. We included indicators for whether

[5]Most recent Centers for Disease Control data indicate the most popular months are July, August, and September. (see https://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58_25.htm).

the official replied and whether the reply included any variation of the word *congratulations*. Neither measure is subject to researcher discretion. The former is a blunt measure of response quality, whereas we argue the latter is a rough proxy for the officials' attitudes related to the couple in question. Other audit studies use salutations to measure differences across treatments (e.g., Einstein and Glick 2017). In this case, responding with "congratulations" is consistent with American norms in conversations about marriage. Deviation from that norm, then, potentially provides a measure of officials' attitudes that may be independent from providing factual responses to questions.

Not all responses were of equal quality. Some provide complete answers to every question, whereas others ask the inquirer to find the information on a webpage (without providing the link to that webpage). Some provide information over and above what was specifically asked about, and others provide no information at all. To measure this quality, we coded responses for whether they contained responses to the three questions posed in Figure 2: how much it cost, how long it was valid, and when it could be obtained. Again, there is wide variation by state, county, and city in each of these questions. The lowest cost was \$4, and the highest was \$200. Licenses are valid anywhere from 2 weeks to, in the words of one official, "eternity." Some officials require appointments, keep different office hours, or reside in states with variable waiting periods. Some webpages did not contain the necessary information to gauge accuracy. Some respondents indicated that their webpages were out of date. Therefore, we simply code whether each question was answered (not whether it was "accurate"). To check the robustness of congratulatory language, we also had coders determine whether the reply included *any* felicitations. This included both congratulatory language and any other well wishes specific to the marriage or wedding.

We used single-blind coding to assess whether the response met the above criteria. Specifically, we removed names of fictitious emailers and all partner synonyms so that coders did not know the treatment condition. Responses were coded by two coders; when coders disagreed, a third resolved the discrepancy. Under this coding scheme, emails that were coded as poor in quality tended to only ask follow-up information, forward the emailer to a webpage, omit answers, and exclude any indication of nuptial excitement. Additional details about this process, along with inter-rater reliability information, can be found in SI Appendixes B and C.

Findings

Overall, we find limited evidence of systematic differences across couples, and no evidence of conditional treatment effects by selection or state legal history. Table 2 reports summary statistics by treatment group. The overall response rate of 71% is comparable to other surveys of local-level officials in the United States (Dynes, Hassell, and Miles 2018; Giulietti, Tonin, and Vlassopoulos 2019; White, Nathan, and Faller 2015) and Germany (Grohs, Adam, and Knill 2015), and higher than those targeting elected officials (Butler and Broockman 2011) and bureaucrats administering federal programs (Einstein and Glick 2017).[6] To avoid posttreatment bias, all response quality measures (including congratulations rates) redefine both nonresponse and nonquality as zeroes (Coppock 2018). For example, a quality response to the cost question ($Cost = 1$) tells the fictitious emailer about fees associated with obtaining a marriage license, whereas for nonquality outcomes ($Cost = 0$), either the official responded without the cost information or no response was received. By implication, overall response rates cannot be less than any response quality rate.

Among measures of quality, information about cost was provided most readily by officials. Congratulatory language was the least common—but still present in about 1 out of every 6 responses. As is typical in email audit studies, errors and bounces were unevenly distributed across treatment groups, which led to imbalance across one or more covariates. In this case, there were significant differences in population across treatment groups (see SI Table A1). Given this, we adopt a parametric approach to estimating treatment effects.

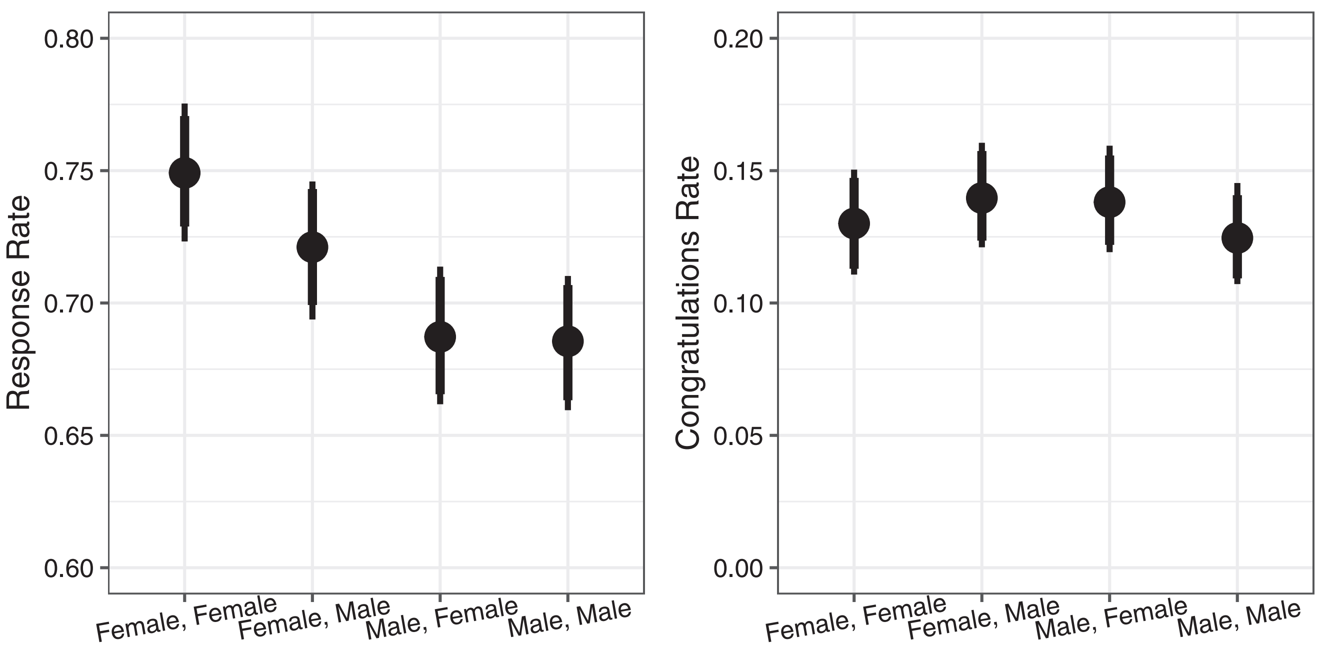
Figure 3 plots predicted probabilities for response and congratulatory language by treatment group with 90% and 95% confidence intervals. Models include treatment group, state legal history, service level, population, selection mechanism, Democratic presidential vote share in jurisdiction, and contact method. We adopt the observed case approach recommended by Hanmer and Ozan Kalkan (2013) to estimate each quantity. Specifically, we estimate first differences after simulating a change in each treatment condition from 0 to 1 for all observations in the data and taking the difference between same-sex and different-sex conditions, while fixing each observation to the observed covariate values. In general, our substantive findings

[6] For a more systematic meta-analysis, see Costa (2017).

TABLE 2 Mean Response by Treatments in All States

Treatment	Response	Congrats	Cost	Valid	When	Well Wishes	N
Male Emailer, Male Spouse	0.678	0.124	0.482	0.469	0.334	0.127	1,067
Female Emailer, Female Spouse	0.744	0.132	0.538	0.510	0.387	0.132	1,094
Male Emailer, Female Spouse	0.684	0.138	0.482	0.468	0.355	0.134	1,088
Female Emailer, Male Spouse	0.718	0.142	0.534	0.503	0.384	0.137	1,072

FIGURE 3 Predicted Probability of Response and Congratulations by Treatment Group



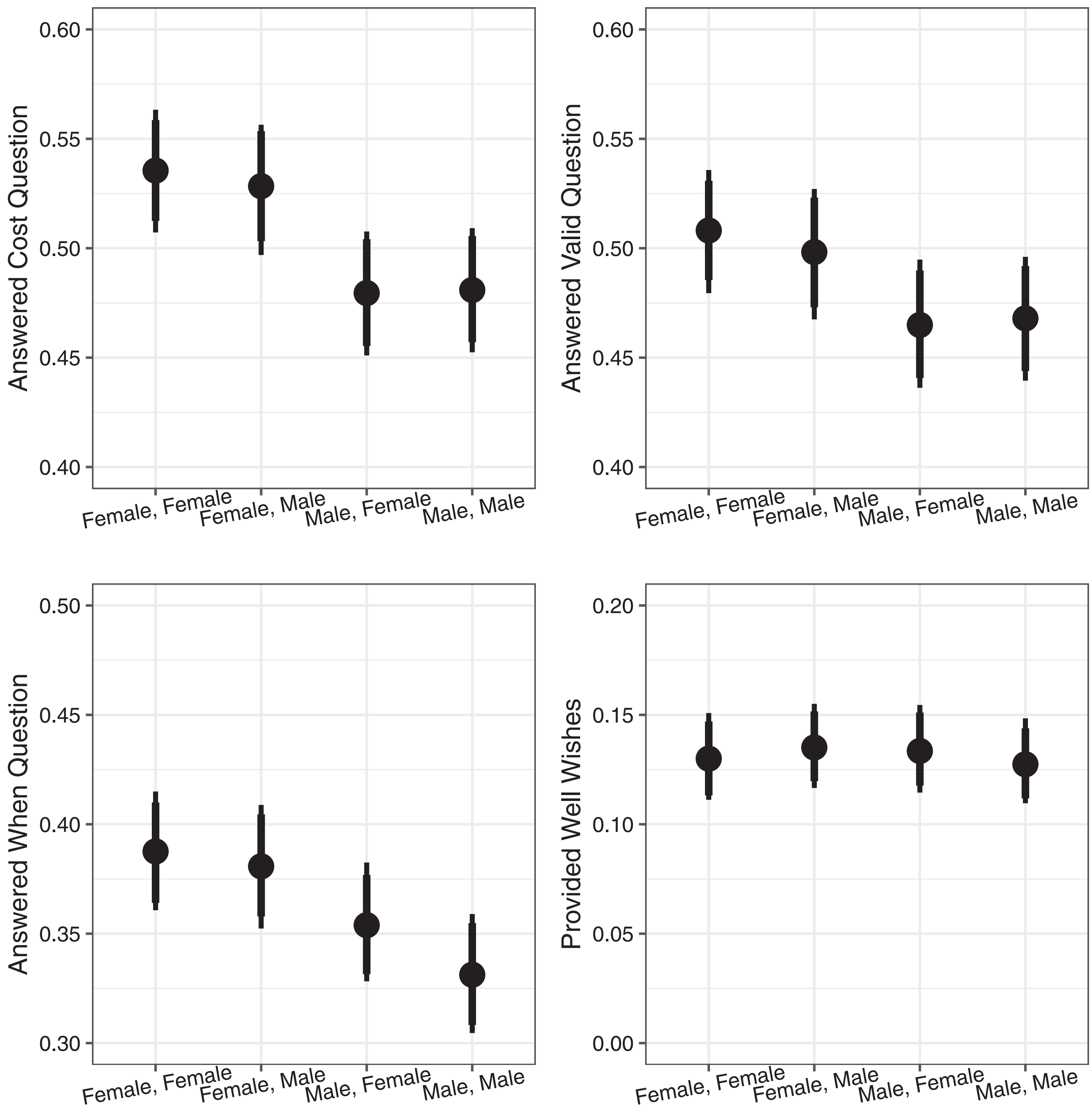
Note: The figure plots point estimates and 90% and 95% confidence intervals simulated from observed data and logistic regression, with controls for state legal history, institution level, population, selection method, democratic presidential vote share, and contact method.

are not sensitive to the covariates included in these models.

Consistent with mean response rates, there are only minor differences across sexual orientation of couples. Officials appeared most likely to respond to female emailers. In general, estimated response rates suggest that number and prominence of female names were more predictive of responsiveness than the sexual orientation of the couple. The difference between same-sex and opposite-sex couples is small in magnitude and not distinguishable from zero (see Figure 7). Moreover, because of the high response rate in the lesbian couple treatment, it runs counter to expectations. Congratulations rates are signed consistent with expectations, but differences for this outcome are, again, small and not distinguishable from zero.

As Figures 4 and 5 suggest, these findings are also consistent across various measures of response quality. Answers to questions about license cost, expiration, and application time were all near zero, though inconsistently signed. The lack of differences in congratulations rates is robust to our hand-coded measure of officials' felicitations. Notably, each quality measure preserves the rank order of responsiveness among treatment groups associated mostly with gender, not sexual orientation. The substantively small and inconsistent direction of the differences in Figure 5 suggests that the results are not an artifact of the particular responsiveness measure. Though we prefer presenting more substantively meaningful outcomes, we should also note there are no differences among aggregated indexes of responsiveness constructed from these indicators.

FIGURE 4 Predicted Probability of Quality Response by Treatment Group

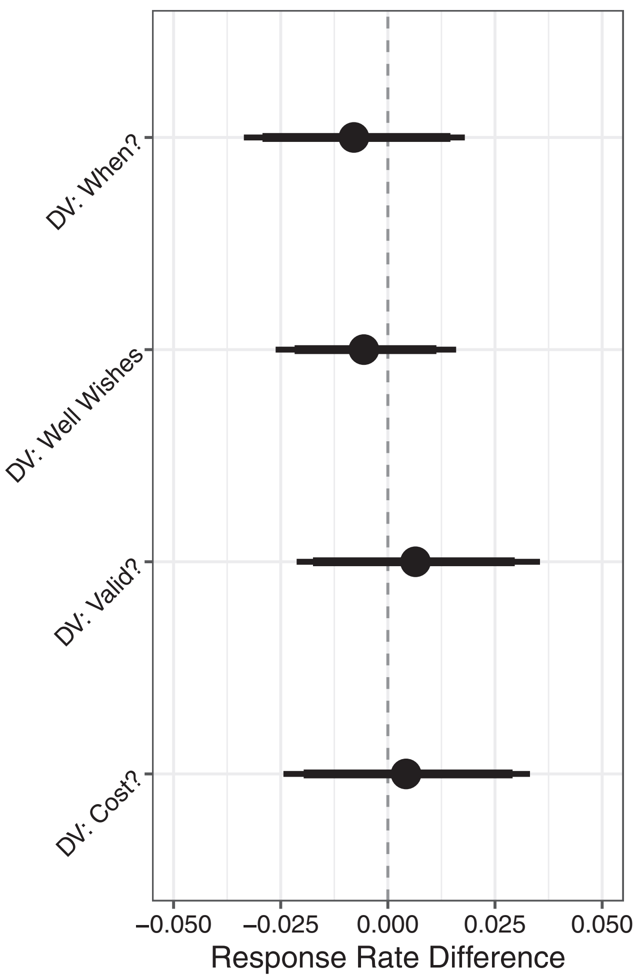


Note: The figure plots point estimates and 90% and 95% confidence intervals simulated from observed data and logistic regression, with controls for state legal history, institution level, population, selection method, democratic presidential vote share, and contact method.

Selection Mechanism and Legal History

Though we are less confident in ruling out small to moderate effects, it is important to note that the results were not conditional on either selection mechanism or state legal history. In fact, most of the overall estimates are similar. We estimated differences among treatments for the 26% of all appointed officials (conditional on local ideology) and those who were not. We also estimated first differences among states with prior legal history unfavorable and favorable to same-sex couples.

FIGURE 5 No Evidence of Differences in the Quality of Responses (Simulated)



Note: Plots the marginal change in predicted probability of response quality with 90% and 95% confidence intervals, where negative values indicate a lower rate for same-sex couples.

Specifically, we group officials in states that had unfavorable laws prior to the Supreme Court decisions in *Lawrence v. Texas* (539 U.S. 558, 2003) and *Obergefell* that overturned them. As Figure 6 suggests, these states overlap but are not synonymous.[7] In general, the conditional average treatment effects (CATEs) we present are not sensitive to the particular measure of state legal history used.

Figure 7 plots differences across states. Not surprisingly, no difference-in-differences is distinguishable from zero. This consistency is also robust to another potential proxy for attitudes: jurisdiction liberalism measured with Democratic presidential vote share. We plot results for CATEs for state legal history, but the same consistency extends to selection mechanism. According to our data, the differences-in-differences among appointed and elected officials by jurisdiction conservatism are small and not distinguishable (by conventional thresholds) from zero. There is no evidence of the selection hypothesis in Figure 8, which plots the interaction of selection mechanism and local-level ideology. Responsiveness differences across putative couples also did not vary by local-level conservatism, independent of selection mechanism. This is notable since partisan differences in support for same-sex marriage remain, and local-level ideology is likely correlated with the preferences of officials.[8] These results are most consistent with White, Nathan, and Faller (2015), and they differ from other studies that demonstrate preferential treatment by partisanship (e.g., Butler and Broockman 2011; Porter and Rogowski 2018).

Though we report differences among outcomes coded automatically for these hypotheses, the results hold for other measures of the quality of responses. Finally, the results are also substantively similar for simple difference-of-means tests. Put simply, we find no evidence of substantial differences across sexual orientation in these marriage license inquiries—regardless of jurisdictional subset, local institutions, or measure of responsiveness.

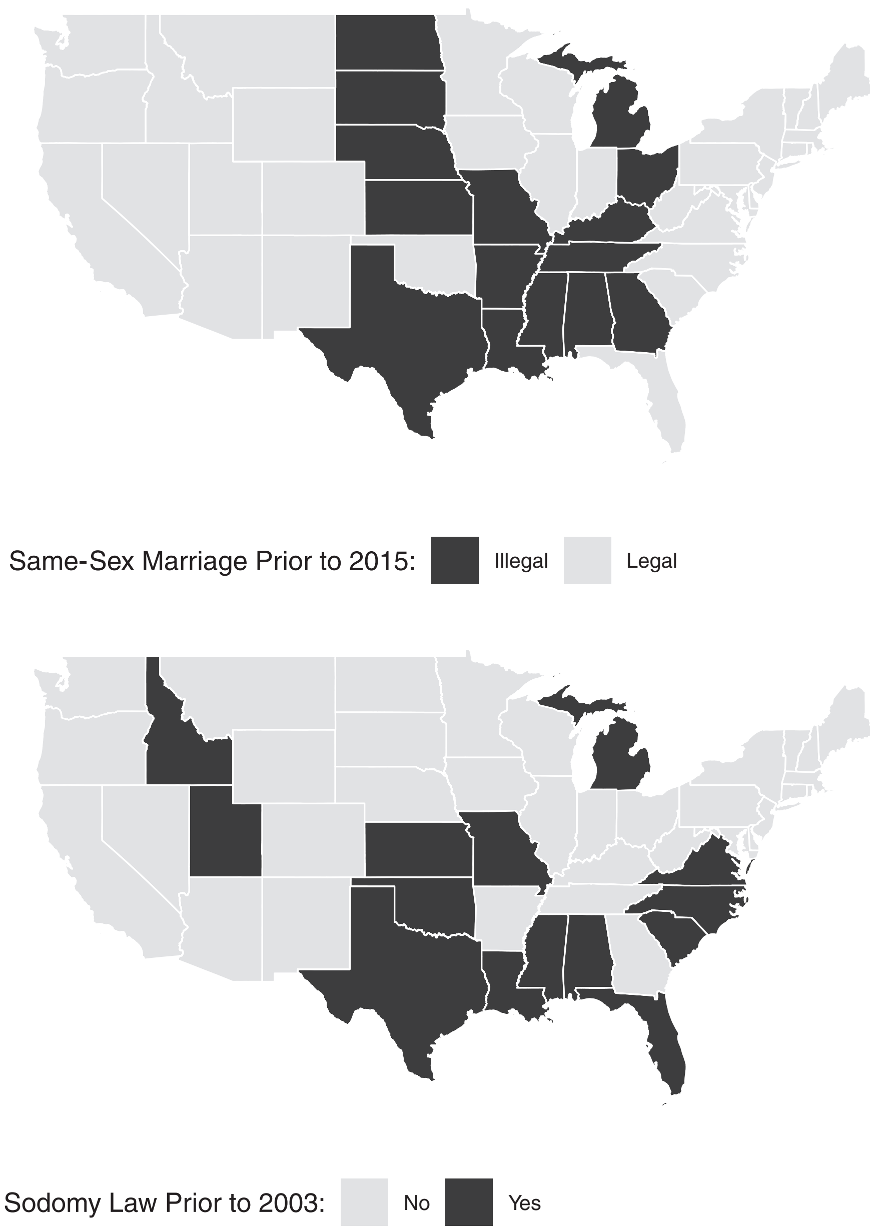
Discussion

The experiment did not uncover differences across couple type that were consistent with our theoretical expectations related to LGBT politics and executive accountability. Our preferred substantive interpretation of this finding is that anecdotes of LGBT couples facing discrimination are extreme cases, and that judicial change was largely successful at achieving bureaucratic compliance. In general, our experiment suggests that differences in (email) correspondence are either small, undetectable by our measures, or just as likely due to chance. We discuss these conclusions in light of several potential concerns about the study design.

[7]These legal histories are marginally correlated at 0.26 ($p < .1$). It is also worth noting that neither legal history appears correlated with selection method, which suggests our design can plausibly separate the two.

[8]One limitation, of course, is that we do not have information on the partisanship of the officials themselves. Although this information is difficult to obtain for many small jurisdictions, it may be worthwhile for future studies.

FIGURE 6 State Laws Affecting Same-Sex Couples

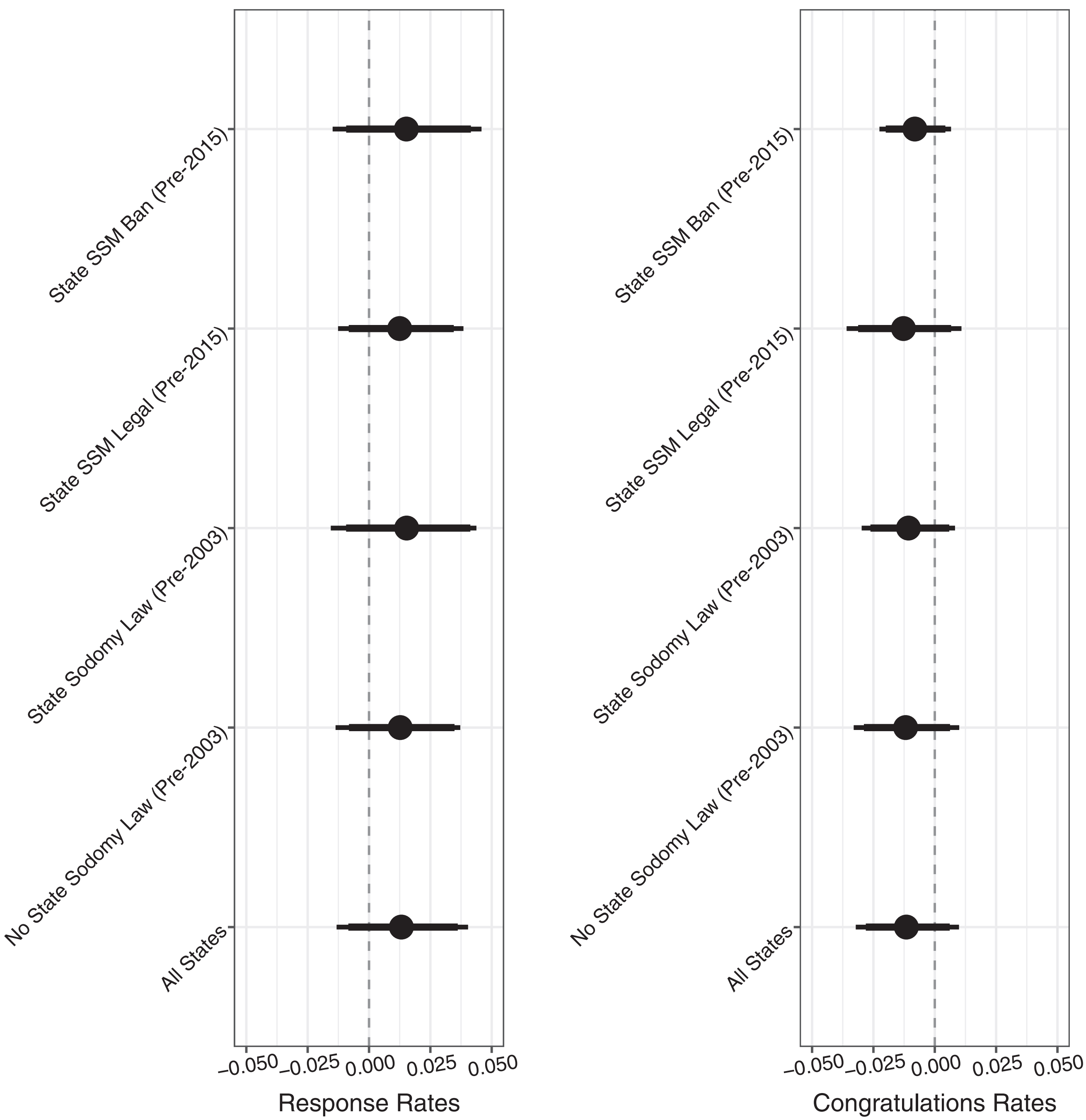


Power

Our sample size is comparable to or exceeds other recent audit studies that find evidence of bias and heterogeneous treatment effects (e.g., Butler and Broockman 2011; Carnes and Holbein 2018; Einstein and Glick 2017; Jilke, Van Dooren, and Rys 2018; White, Nathan, and Faller 2015). However, since our results suggest no bias, it is important to contextualize the magnitude of these effects and the power of the design. We argue that the interpretation that there is no evidence of moderate to large bias is warranted for several reasons.

First, the conventional power threshold of 80% suggests that our study should uncover effect sizes of 4 percentage point differences (or more) in responsiveness (see SI Figure A1). Importantly, other email audit studies have found more substantial effects for bias against other marginalized communities. For Latino names, for example, White, Nathan, and Faller (2015) find a reduction in responsiveness of 5–7 percentage points, and Einstein and Glick (2017) find differences in email tone of 20 percentage points. Thus, one way of contextualizing our findings is that they suggest bias against LGBT people is likely less than what is typically found for

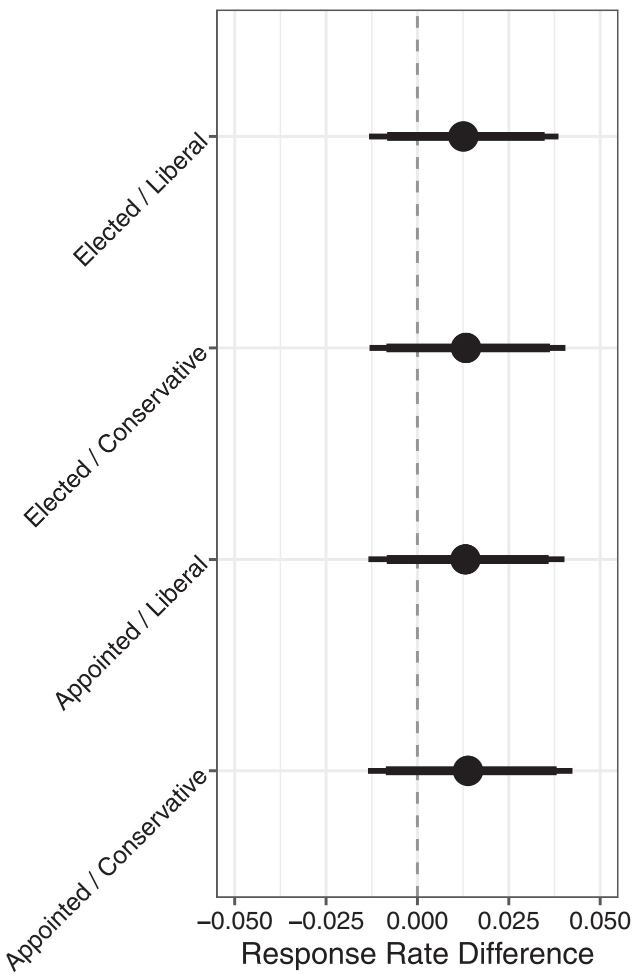
FIGURE 7 No Evidence of Responsiveness Differential from Simulated Differences



Note: Plots the marginal change in predicted probability of response or congratulations with 90% and 95% confidence intervals, where negative values indicate a lower rate for same-sex couples. "All States" corresponds to the experiment average treatment effect; all other estimates are CATE based on state subsets.

other marginalized groups. Second, and relatedly, we are more confident in these conclusions because our survey instrument generated multiple outcomes that point to the same basic finding. Finally, we find substantial differences by gender, which seem to countervail the hypothesis that sexual orientation drives bias. Couples with two female names consistently exhibited the highest levels of responsiveness.

FIGURE 8 No Evidence of Differences in Response Rates (Simulated) by Institution and Ideology



Note: Plots the marginal change in predicted probability of response quality with 90% and 95% confidence intervals, where negative values indicate a lower rate for same-sex couples. For the purposes of this figure, liberal and conservative are defined as 35% and 65% Democratic two-party vote share in the 2008 and 2012 presidential elections.

Missing Officials

There is no national registry or list of marriage license–granting officials. Our best estimate is 5,123. We arrived at this list by using census lists of counties in states where officials operated at that jurisdiction, and municipalities in states in which licenses are granted by cities. However, some city officials in small municipalities indicated they had functionally delegated this role to larger cities nearby. In one case, an official indicated she had not received a marriage license application "in ages." Outdated city population estimates suggest some municipalities legally permitted to issue licenses do not because there is no official employed to do the job. Other jurisdictions may be counted as cities by the census but are now unincorporated territory.[9] Some jurisdictions' webpages indicate they issue licenses, but they provide only a physical address or phone contact information. Other jurisdictions had no web presence. These officials were excluded and tended to be in smaller, more rural, and politically conservative localities.

Thus, we also reestimated treatment effects using inverse probability score weighting. In this analysis, the first-stage predictors were population, local ideology, state legal history, selection method, and official level. As SI Figure A3 indicates, our results are robust to this correction. This is not especially surprising since we found no heterogeneous treatment effects based on these pretreatment covariates.

But this cannot rule out the possibility of an unobserved factor orthogonal to local ideology (or another covariate we observe) that predicts both discriminatory behavior and the availability of contact information. The substantive takeaway is that our analysis cannot speak to the behavior of these omitted officials. If discriminatory behavior was sufficiently prevalent within this relatively small subset of cases, it could change our conclusions. Although it is beyond the scope of this study, future research might address this limitation by conducting an audit via telephone.

Treatment

Our treatment was designed to closely resemble a typical email inquiry officials would receive. It requires the official to make an inference about sexual orientation based on names and a gendered partner synonym. One potential explanation for our finding is that the content of the treatment was simply missed. But the subtlety of this treatment mirrors that of numerous email audit studies that uncover evidence of bias. There is also evidence that the names in the outgoing emails were read. Roughly 43% of responses used the first name of the fictitious emailer in their reply, a lower bound for those who read the names that contained the treatment. One clerk even signaled support for marriage equality in a state that previously had banned same-sex marriage. After providing all information and congratulating the couple by name, the clerk wrote: "...Let me say that we are here to serve

[9] One example is Centerville, Maine, whose fewer than 50 residents voted to dissolve the town government after the 2000 census.

all [state residents] and we are eager to help you get your marriage license."

The experiment did suggest, however, that for the roughly 1 in 10 offices that serve over 250,000 constituents, the use rate for first names in reply emails was about 3 percentage points lower. This may indicate that offices with high caseloads may be less receptive to treatment. But the small number of these cases, and the weak magnitude of the effect, along with evidence from existing studies, all suggest this is not a sufficient explanation for our findings. This, of course, does not rule out the possibility of systematic differences in face-to-face or phone interactions, which would require far more resources to study properly.

Hawthorne Effect

No email response suggested discovery of the experiment. However, this does not rule out the possibility that officials changed their behavior because they knew they were being observed. The odds of discovery may be increasing as audit studies of officials become more common.

In this study, however, we believe the likelihood of the Hawthorne effect driving our results is minimal. First, we did not contact any officials, beforehand, either with an attempt at a bias-reducing intervention (à la Butler and Crabtree 2017) or to collect pretreatment data. Pretreatment surveys are one noted driver of the Hawthorne effect in clinical trials (e.g., McCambridge and Kypri 2011). Second, to our knowledge, the marriage license–granting officials in our sample have never been subjects in an audit study. We argue this makes discovery less likely among this population, relative to officials like state legislators, who are frequently targeted. Finally, overall responsiveness among these officials was high, relative to other officials. The observed response rate of 71% ranks sixth highest, compared with the 41 studies reviewed by Costa (2017). The Hawthorne effect is typically more pronounced among subjects with lower baseline performance (Granberg and Holmberg 1992; McCambridge, Witton, and Elbourne 2014).

Conclusion

Our study provided an experimental test of bureaucratic responsiveness to same-sex and opposite-sex couples requesting information about obtaining a marriage license. In contrast to other audit studies investigating discrimination against marginalized groups, we find little to no evidence of this toward LGBT couples seeking a marriage license. Our findings have important implications for research on LGBT politics and executive accountability, more broadly.

First, our study examines the implications of a policy recently settled by the Supreme Court. Historical moments in which bureaucrats must implement far-reaching policy change are necessarily rare, which sets our analysis apart from other audit studies. Our null results coupled with the timing of our study, just 3 years after 14 states were forced to legalize same-sex marriage, suggests that executives, who have discretion to implement law, have largely complied with judicial directives related to same-sex marriage. This contribution is best situated in broader historical context surrounding the implementation of civil rights for minority groups. In other cases of civil rights, bureaucratic officials slowed the implementation of civil rights for African American and underrepresented minorities. When viewed alongside these other civil rights examples, our findings suggest that discriminatory biases toward marginalized groups vary across group and policy context.

We are not suggesting that LGBT people no longer face discrimination in the United States. LGBT people continue to report experiences of discrimination at high rates (Pew Research Center 2013). In 2017, the Human Rights Campaign documented 129 anti-LGBT bills introduced in states (Warbelow and Diaz 2017). President Trump has implemented a ban on transgender people's service in the military (Marimow 2019). Recent work by Sun and Gao (2019) suggests lending practices disadvantage same-sex couples. Moreover, our findings do not speak to more subtle bias not typically detectable in a study of this size, and it is important to note that our evidence is confined to how public officials respond to putative constituents. At a minimum, however, our study aligns with other other evidence that shows improving social and political acceptance of LGBT people.

This raises the question of what may account for differences in observed patterns of discrimination across race, ethnicity, and sexuality. First, it is possible the comparatively apolitical character of the officials and task suppress potential bias. Notably, one of the few studies to find mixed evidence of racial discrimination was conducted on street-level bureaucrats administering housing policy (Einstein and Glick 2017). Marriage license–granting officials are more similar to these bureaucrats than to officials situated in positions tied directly to the politics of voting administration (e.g., Butler and Broockman 2011; White, Nathan, and Faller 2015). Our study is also consistent with emerging research that suggests bureaucratic agents in some policy contexts that are concerned chiefly with service provision exhibit little evidence of bias toward putative constituents (e.g., Einstein and Glick 2017; Jilke, Dooren, and Rys 2018; Porter and Rogowski 2018). These areas differ relative to other contexts, such as policing and government assistance programs, where researchers have documented racial bias.

Second, the recipients of social assistance are often constructed as undeserving and in racial terms (e.g., Gilens 1999; Schneider and Ingram 1993), which increases the likelihood of discrimination, even among those who are less tied to politics. Notably, our findings stand in contrast to evidence of discrimination in another area of service provision—federal assistance programs (Keiser, Mueser, and Choi 2004; Schram et al. 2009). In contrast, marriage licenses are not redistributive. Moreover, marriage constitutes a form of assimilation into mainstream institutions and may privilege advantaged subgroups in the LGBT community (e.g., Cohen 1997; Strolovitch 2012). Thus, discrimination on the basis of sexual orientation may be less likely in this policy context, relative to those involving both elected officials, race, or the distribution of resources.

Our study also suggests avenues for future research. For scholars of LGBT politics, the difference in bureaucratic responsiveness to gay and lesbian couples is particularly noteworthy. This finding is similar to research that has found gay men receive lower evaluations compared to lesbians in other contexts, and more work needs to examine the foundations of these gender- and sexuality-based biases. Importantly, this is largely driven by the finding that bureaucrats were more responsive to female requesters in general. Since we did not have initial expectations that our results would be driven by requester gender, we do not engage in post hoc analyses. One possibility, however, is that women may be more likely to serve in public office as license-issuing authorities, and we could be observing higher responsiveness to other women. This would be consistent with findings of responsiveness to in-group constituents (Butler and Broockman 2011). Future research can address this topic, contributing to a growing literature about the representation of women in local government (Einstein, Palmer, and Glick 2018; Holman 2017). Finally, our findings are limited to understanding discrimination against white LGBT people, and future research should explicitly examine discrimination and biases at the intersection of race, sexuality, and gender. Such an agenda might help scholars better understand the complexities and contingencies of discrimination in the United States.

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016 . "The Political Legacy of American Slavery." *Journal of Politics* 78(3): 621–41.

Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." *American Economic Review* 85(3): 304–21.

Badgett, M. V. Lee, Brad Sears, Holning Lau, and Deborah Ho. 2008. "Bias in the Workplace: Consistent Evidence of Sexual Orientation and Gender Identity Discrimination 1998–2008." *Chicago-Kent Law Review* 84(2): 559–95.

Bailey, John, Michael Wallace, and Bradley Wright. 2013. "Are Gay Men and Lesbians Discriminated Against When Applying for Jobs? A Four-City, Internet-Based Field Experiment." *Journal of Homosexuality* 60(6): 873–94.

Bertrand, Marianne, and Sendhil Mullainathan. 2003. "Are Emily and Greg More Employable Than Lakisha and Jamal?" NBER Working Paper No. 9873. https://www.nber.org/papers/w9873.pdf.

Besley, Timothy, and Stephen Coate. 2003. "Elected versus Appointed Regulators: Theory and Evidence." *Journal of the European Economic Association* 1(5): 1176–1206.

Blinder, Alan, and Richard Perez-Pena. 2015. "Kentucky Clerk Denies Same-Sex Marriage Licenses, Defying Court." *New York Times*, September 1. https://www.nytimes.com/2015/09/02/us/same-sex-marriage-kentucky-kim-davis.html.

Butler, Daniel M., and David E. Broockman. 2011. "Do Politicians Racially Discriminate against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55(3): 463–77.

Butler, Daniel M., and Charles Crabtree. 2017. "Moving Beyond Measurement: Adapting Audit Studies to Test Bias-Reducing Interventions." *Journal of Experimental Political Science* 4(1): 57–67.

Canes-Wrone, Brandice, Tom S. Clark, and Jason P. Kelly. 2014. "Judicial Selection and Death Penalty Decisions." *American Political Science Review* 108(1): 23–39.

Carnes, Nicholas, and John B. Holbein. 2018. "Do Public Officials Exhibit Social Class Biases When They Handle Casework? Evidence from Multiple Correspondence Experiments." Working paper. https://bit.ly/2GN29gb.

Cohen, Cathy J. 1997. "Straight Gay Politics." In *Ethnicity and Group Rights*, ed. Ian Shapiro and William Kymlicka. New York: New York University Press, 572–616.

Coppock, Alexander. 2018. "Avoiding Post-Treatment Bias in Audit Experiments." *Journal of Experimental Political Science* (May): 10–13.

Costa, Mia. 2017. "How Responsive Are Political Elites? A Meta-Analysis of Experiments on Public Officials." *Journal of Experimental Political Science* 4(3): 241–54.

Crabtree, Charles, and Volha Chykina. 2018. "Last Name Selection in Audit Studies." *Sociological Science* 5(1): 21–28.

Crow, Stephen M., Lillian Y. Fok, Sandra J. Hartman, and Dinah M. Payne. 1991. "Gender and Values: What Is the Impact on Decision Making?" *Sex Roles* 25(3–4): 255–68.

Daugherity, Brian J., and Charles C. Bolton. 2008. *With All Deliberate Speed: Implementing Brown v. Board of Education*. Fayetteville: University of Arkansas Press.

Distelhorst, Greg, and Yue Hou. 2014. "Ingroup Bias in Official Behavior: A National Field Experiment in China." *Quarterly Journal of Political Science* 9: 203–30.

Doleac, Jennifer L., and Luke C. D. Stein. 2013. "The Visible Hand: Race and Online Market Outcomes." *Economic Journal* 123(572): 469–92.

Dunigan, Jonece Starr. 2018. "Alabama Sees Hundreds of Same-Sex Marriages Each Year, Despite Eight Holdout Counties." al.com, October 16. https://www.al.com/news/birmingham/2018/10/alabama-sees-hundreds-of-same-sex-marriages-each-year-despite-eight-holdout-counties.html

Dynes, Adam M., Hans J. G. Hassell, and Matthew R. Miles. 2018. "Political Ambition and Constituent Service: Does Ambition Influence How Local Officials Respond to Electoral and Non-Electoral Service Requests?" Working Paper. https://bit.ly/2tAfY9t.

Einstein, Katherine Levine, and David M, Glick. 2017. "Does Race Affect Access to Government Services? An Experiment Exploring Street-Level Bureaucrats and Access to Public Housing." *American Journal of Political Science* 61(1): 100–116.

Einstein, Katherine Levine, Maxwell Palmer, and David M, Glick. 2018. "Who Participates in Local Government? Evidence from Meeting Minutes." *Perspectives on Politics* 17(1): 28–46.

Fields, Josepha, Linda Klein, and James Sfiridis. 1997. "A Market-Based Evaluation of the Election versus Appointment of Regulatory Commissioners." *Public Choice* 92(3): 337–51.

Gell-Redman, Micah, Neil Visalvanich, Charles Crabtree, and Christopher J, Fariss. 2018. "It's All about Race: How State Legislators Respond to Immigrant Constituents." *Political Research Quarterly* 71(3): 517–31.

Gilens, Martin. 1999. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. Chicago: University of Chicago Press.

Giulietti, Corrado, Mirco Tonin, and Michael Vlassopoulos. 2019. "Racial Discrimination in Local Public Services: A Field Experiment in the United States." *Journal of the European Economic Association* 17: 165–204. https://academic.oup.com/jeea/article/17/1/165/4756072?guestAccessKey=e19c9228-8699-4f93-812e-4d99a35e5252

Granberg, Donald, and Soren Holmberg. 1992. "The Hawthorne Effect in Election Studies: The Impact of Survey Participation on Voting." *British Journal of Political Science* 22(2): 240–47.

Gray, Thomas. 2017. "The Influence of Legislative Reappointment on State Supreme Court Decision-Making." *State Politics and Policy Quarterly* 17(3): 275–98.

Gray, Thomas. 2019. "Strategic Deference in State Supreme Courts: Executive Influence on Judicial Decision-Making in Executive Reappointment States." *Journal of Law, Economics, and Organization* 35(2): 422–53.

Grohs, Stephan, Christian Adam, and Christoph Knill. 2015. "Are Some Citizens More Equal Than Others? Evidence from a Field Experiment." *Public Administration Review* 76(1): 155–64.

Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57(1): 263–77.

Hebl, Michelle R., Jessica Bigazzi Foster, Laura M. Mannix, and John F. Dovidio. 2002. "Formal and Interpersonal Discrimination: A Field Study of Bias Toward Homosexual Applicants." *Personality and Social Psychology Bulletin* 28(6): 815–25.

Holman, Mirya R. 2017. "Women in Local Government." *State and Local Government Review* 49(4): 285–96.

Huber, Gregory A., and Sanford C. Gordon. 2004. "Accountability and Coercion: Is Justice Blind When It Runs for Office?" *American Journal of Political Science* 48(2): 247–63.

Jilke, Sebastian, Wouter Van Dooren, and Sabine Rys. 2018. "Discrimination and Administrative Burden in Public Service Markets: Does a Public Private Difference Exist?" *Journal of Public Administration Research and Theory* 28(3): 423–39.

Jones, David A. 1996. "Discrimination against Same-Sex Couples in Hotel Reservation Policies." *Journal of Homosexuality* 31(1–2): 9–41.

Keiser, Lael R., Peter R. Mueser, and Seung-Whan Choi. 2004. "Race, Bureaucratic Discretion, and the Implementation of Welfare Reform." *American Journal of Political Science* 48(2): 314–27.

Kimball, David C., and Martha Kropf. 2006. "The Street-Level Bureaucrats of Elections: Selection Methods for Local Election Officials." *Review of Policy Research* 23(6): 1257–68.

Lim, Claire S. H. 2013. "Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges." *American Economic Review* 103(4): 1360–97.

Marimow, Ann E. 2019. "Restriction on Transgender Troops Serving in Military Can Stand for Now, D.C. Federal Appeals Court Rules." *Washington Post*, January 4.

Marschall, Melissa J., and Amanda Rutherford. 2016. "Voting Rights for Whom? Examining the Effects of the Voting Rights Act on Latino Political Incorporation." *American Journal of Political Science* 60(3): 590–606.

Maynard-Moody, Steven, and Michael C, Musheno. 2003. *Cops, Teachers, Counselors: Stories from the Front Lines of Public Service*. Ann Arbor: University of Michigan Press.

McCambridge, Jim, and Kypros Kypri. 2011. "Can Simply Answering Research Questions Change Behaviour? Systematic Review and Meta Analyses of Brief Alcohol Intervention Trials." *PLOS ONE* 6(10): 1–9.

McCambridge, Jim, John Witton, and Diana R. Elbourne. 2014. "Systematic Review of the Hawthorne Effect: New Concepts Are Needed to Study Research Participation Effects." *Journal of Clinical Epidemiology* 67(3): 267–77.

Mendez, Matthew S., and Christian R. Grose. 2018. "Doubling Down: Inequality in Responsiveness and the Policy Preferences of Elected Officials." *Legislative Studies Quarterly* 43(3): 457–91.

Milkman, Katherine L., Modupe Akinola, and Dolly Chugh. 2014. "What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway into Organizations." https://www.apa.org/pubs/journals/releases/apl-0000022.pdf.

Minta, Michael D. 2011. *Oversight: Representing the Interests of Blacks and Latinos in Congress.* Princeton, NJ: Princeton University Press.

Moore, Ryan T. 2012. "Multivariate Continuous Blocking to Improve Political Science Experiments." *Political Analysis* 20(4): 460–79.

Moore, Ryan T., and Keith Schnakenberg. 2016. "blockTools." R Package 0.6-3 https://CRAN.R-project.org/package=blockTools.

Moynihan, Donald P., Pamela Herd, and Hope Harvey. 2015. "Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions." *Journal of Public Administration Research and Theory* 25(1): 43–69.

Nunn, Nathan. 2009. "The Importance of History for Economic Development." *Annual Review of Economics* 1: 65–92.

Pager, Devah, Bart Bonikowski, and Bruce Western. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74: 777–99.

Pepinsky, Thomas B., Jan H. Pierskalla, and Audrey Sacks. 2017. "Bureaucracy and Service Delivery." *Annual Review of Political Science* 20: 249–68.

Perry, James L. 1996. "Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity." *Journal of Public Administration Research and Theory* 6(1): 5–22.

Pew Research Center. 2013. *A Survey of LGBT Americans: Attitudes, Experiences and Values in Changing Times.* Washington, DC: Pew Research Center.

Pew Research Center. 2017. "Changing Attitudes on Gay Marriage." Pew Forum on Religion and Public Life, 1–8. https://www.pewforum.org/fact-sheet/changing-attitudes-on-gay-marriage/.

Porter, Ethan, and Jon C, Rogowski. 2018. "Partisanship, Bureaucratic Responsiveness, and Election Administration: Evidence from a Field Experiment." *Journal of Public Administration Research and Theory* 28(4): 602–17.

Prager, Tyler. 2018. "Cuomo Opens Investigation into Denial of Marriage License to Gay Couple." *New York Times*, August 2. https://www.nytimes.com/2018/08/02/nyregion/cuomo-same-sex-marriage-clerk.html

Robertson, Campbell. 2016. "Roy Moore, Alabama Chief Justice, Suspended over Gay Marriage Order." *New York Times*, September 30. https://www.nytimes.com/2016/10/01/us/roy-moore-alabama-chief-justice.html.

Sanchez, Sam. 2018. "A County Clerk in Texas Still Won't Sign Same-Sex Marriage Licenses." *San Antonio Current*, August 6. https://www.sacurrent.com/the-daily/archives/2018/08/06/a-county-clerk-in-texas-still-wont-sign-same-sex-marriage-licenses.

Schneider, Anne, and Helen Ingram. 1993. "Social Construction of Target Populations: Implications for Politics and Policy." *American Political Science Review* 87(2): 334–47.

Schram, Sanford F., Joe Soss, Richard C. Fording, and Linda Houser. 2009. "Deciding to Discipline: Race, Choice, and Punishment at the Frontlines of Welfare." *American Sociological Review* 74(3): 398–422.

Spriggs, James F. 1996. "The Supreme Court and Federal Administrative Agencies: A Resource-Based Theory and Analysis of Judicial Impact." *American Journal of Political Science* 40(4): 1122–51.

Strolovitch, Dara Z. 2012. "Intersectionality in Time: Sexuality and the Shifting Boundaries of Intersectional Marginalization." *Politics & Gender* 8(3): 386–97.

Sun, Hua, and Lei Gao. 2019. "Lending Practices to Same-Sex Borrowers." *Proceedings of the National Academy of Sciences* 116(19): 9293–9302.

Trounstine, Jessica. 2018. *Segregation by Design: Local Politics and Inequality in American Cities.* New York: Cambridge University Press.

Turner, Margery Austin, Rob Santos, Diane K, Levy, Doug Wissoker, Claudia Aranda and Rob Pitingolo. 2013. *Housing Discrimination against Racial and Ethnic Minorities.* Washington, DC: U.S. Department of Housing and Urban Development.

Van Hoye, Greet, and Filip Lievens. 2003. "The Effects of Sexual Orientation on Hirability Ratings: An Experimental Study." *Journal of Business and Psychology* 18(1): 15–30.

Walters, Andrew S., and Maria-Cristina Curran. 1996. "'Excuse Me, Sir? May I Help You and Your Boyfriend?.'" *Journal of Homosexuality* 31(1–2): 9–41.

Warbelow, Sarah, and Breanna Diaz. 2017. "State Equality Index: A Review of State Legislation Affecting the Lesbian, Gay, Bisexual, Transgender and Queer Community and a Look Ahead in 2018." https://www.hrc.org/blog/hrc-releases-2017-state-equality-index.

Weichselbaumer, Doris. 2003. "Sexual Orientation Discrimination in Hiring." *Labour Economics* 10(6): 629–42.

White, Ariel R., Noah L. Nathan, and Julie K. Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109(1): 129–42.

Wood, B. Dan, and Richard W. Waterman. 1994. *Bureaucratic Dynamics: The Role of Bureaucracy in a Democracy.* Boulder, CO: Westview Press.

Wooston, Cleve R., Jr., and Sandhya Somashekhar. 2017. "A Lesbian Couple Got Their Marriage License—After a County Clerk Called Them an Abomination.'" *Washington Post*, April 19. https://www.washingtonpost.com/news/post-nation/wp/2017/04/19/a-lesbian-couple-got-their-marriage-license-after-a-w-va-clerk-called-them-an-abomination/.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: Study Diagnostics
Appendix B: Response Coding Procedures
Appendix C: Estimation Results