Supplementary Materials

CONTENTS

Appendices **1**

A Contribution to the Literature on Violence against Women in Politics **2**
A.1 Theoretical Contribution . . . 3
A.2 Conceptualizing and Operationalizing Violence . . . 5
A.3 Empirical Contribution . . . 6
A.4 Implications for the Future Study of VAWIP . . . 7
A.5 Supplemental References . . . 8

B Detailed Methods and Materials **13**
B.1 Sampling Frame . . . 13
B.2 Treatments . . . 13
B.3 Messages . . . 13
B.3.1 Study 1 . . . 13
B.3.2 Study 2 . . . 13
B.3.3 Number of Texts Sent . . . 13
B.4 Dependent Variables . . . 14
B.4.1 Coding Instructions . . . 15
B.4.2 Alternative Approaches to Measurement . . . 15
B.5 Robustness of Offensiveness Codings . . . 17
B.5.1 Study 2 - Ordinal measure of offensiveness . . . 17
B.5.2 Different ways of dichotomizing offensiveness . . . 17
B.5.3 Results for Discouragement . . . 18
B.6 Study Context, Delivery, and Timing . . . 19
B.6.1 Supporter Gender . . . 19

C Results and Robustness Checks **20**
C.1 OLS Regressions for Results Reported Visually . . . 20
C.1.1 Offensiveness . . . 20
C.1.2 Silencing . . . 21
C.1.3 Withdrawal . . . 21
C.2 Outcomes within Responses . . . 21
C.2.1 Response Rates . . . 21
C.2.2 Regression Analyses Conditional on Receiving a Response . . . 24
C.2.3 Other Outcomes . . . 27
C.3 Outcomes Moderated by Volunteer Gender . . . 27
C.4 Logistic Regressions . . . 29
C.4.1 Offensiveness . . . 29
C.4.2 Silencing . . . 29
C.4.3 Withdrawal . . . 30
C.5 OLS Regressions Using Cluster-Robust Standard Errors . . . 30
C.5.1 Offensiveness . . . 31
C.5.2 Silencing . . . 32
C.5.3 Withdrawal . . . 32
C.6 Instrumental Variable Regressions . . . 32
C.6.1 Offensiveness . . . 32
C.6.2 Silencing . . . 34
C.6.3 Withdrawal . . . 34

APSR Submission Template

CONTRIBUTION TO THE LITERATURE ON VIOLENCE AGAINST WOMEN IN POLITICS

The existing literature on violence against women in politics is growing rapidly. Because of that, it isn't possible to provide a comprehensive overview of the relevant research in a short-format paper. As such, it is possible that for some readers our paper leaves unanswered what our study contributes to this field. To supplement the shorter literature review presented in the paper, we have conducted a more extensive literature review presented here in the appendix.

We took three sources—Thomas et al. (2019), Krook (2020), and Herrick et al. (2021)—as starting points for an extensive snowball-style review of the VAWIP literature. We provide a list of the resulting 87 books and articles we identified and reviewed at the bottom of this appendix.[1]

We focused our review on three types of sources. First, we looked for theory-oriented pieces, such as Piscopo (2016), Bardall et al. (2020), and Krook's (2020) book, which would give us a rich sense of current debates and developments in the field. Second, we looked for conceptualization- and measurement-oriented pieces, such as Ballington (2018) and Thomas et al. (2019), which would help us situate our concepts and measures within this field. And third, we sought out empirically-oriented pieces that examine violence, harassment, and incivility expressed toward women in politics (e.g., Bjarnegård et al. 2020; Håkansson 2019; Prillaman 2024). This branch of the literature is exceptionally extensive, covering everything from the use of rape as a tool of war to misogynistic tweets toward political activists to targeted killings of politicians. As a result, we paid particular attention to how such violence, harassment, and incivility was expressed through digital fora such as texts, emails, phone calls, social media posts, and written comments. While our resulting list does cover a fair bit of ground, we could not possibly hope to fairly treat such a large literature in a short-format article, and so we have prioritized those pieces that appear to be "near neighbors" of our topic (albeit ones that cover many different countries, political actors, and types of violence).[2]

In addition to surveying the scholarship on VAWIP, we also deepened our literature review in two adjacent fields, incivility (e.g., Cortina 2008; Stryker et al. 2016) and digital misogyny (e.g., Sobieraj 2020). While these literatures were present in the original manuscript, it became clear as we read the VAWIP literature that one important theoretical contribution our paper could make is to more fully integrate these three literatures. While many scholars in these areas cross-cite (e.g., Krook (2020) cites Sobieraj (2020); Sobieraj (2020) cites Cortina et al. (2002); etc.), much of this work currently draws on parallel but separate theories to study the same phenomena. Below, we use the theoretical framework laid out in Krook (2020) to illustrate where overlap and disagreement exists between these literatures and the scholarship on VAWIP, and argue that it would be fruitful for future scholarship that sits at this intersection to pursue a more integrated approach.

Critically, weaving together these literatures makes clearer the importance of recognizing women as representatives (Pitkin 1967; Sobieraj 2020) when they "do" politics—or even are viewed as doing politics, which may be more pertinent—outside the formal spheres of elections and policymaking. We argue that since perpetrators of VAWIP target individual women because they view them to be symbolic and/or substantive representatives of women as a group, scholarship on VAWIP must also consider how women "stand" and "act for" other women in "informal" political contexts—not just in journalism, or NGO work, or even political volunteering as we study here, but on social media and in everyday life as well. Our focus on women as representatives makes it clear that political action (here, as a volunteer) takes place in what was previously understood to be the social or private spheres, and is therefore shaped by both social and political attitudes about gender. To claim politics is or should be separate from the social "sphere" is problematic: we simply can't keep the political separate from the social, either theoretically or empirically, though of course any individual study will focus on particular cases and contexts (e.g., violence against women in elections, or workplaces).

Building on this literature review, we also clarify our concepts and variable labels by situating them in the literature. It was clear from this review that we needed to separate out what had previously been a single

[1] Note that the reference list also includes literature cited in this supplement, like Pitkin (1967), and so adds to more than 87 books and articles.

[2] We also focused on texts that were available in English: this excluded relevant articles like Albaine (2015) and Cerva Cerna (2014) (where only English-language abstracts were available) and would otherwise have been included.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

silencing variable into two separate measures: "withdrawals," which refers to polite or inoffensive requests not to be contacted further (e.g., "please stop"), and "silencing," which now refers only to offensive requests not to be contacted further (e.g., "stfu," which means "shut the fuck up"). We believe the new measures better distinguish theoretically and empirically distinct behaviors.

Finally, our review also revealed more clearly the empirical contribution we make to the scholarship on VAWIP by studying the harassment of political volunteers. While we argued in the original manuscript that these literatures focus heavily on elites—and we still believe this is true—we now understand this to be a function of the nature of the problem and the unique methodological challenges to studying this topic. Because of these challenges, the focus on elites is likely to persist without conscious intervention. We argue that studying less-visible participants in politics is important, both because visibility is itself an important theoretical feature of the VAWIP literature (Thomas et al. 2019; Håkansson 2019), and because of the implications that studying violence against a wider range of political actors has for our understanding of how VAWIP shapes who enters the political arena. If VAWIP prevents women from ever entering politics, scholars focusing exclusively on formal politics will misunderstand the scope and implications of the use of VAWIP.

Theoretical Contribution

We argued above that it is worth more deeply integrating the literatures on VAWIP, gendered incivility, and digital misogyny. We begin by supplying Krook (2020)'s definition of VAWIP: "its central motivation is thus not to gain the upper hand in a game of partisan competition, but rather to exclude women as a group from public life" (4). This makes VAWIP impossible to subsume under narrower frameworks like VAWIE (Krook and Restrepo Sanín 2016b; Krook 2017, 2020,b), which focuses exclusively on women in elections, or existing definitions of political violence, which emphasize physical forms of violence and connote ideological or instrumental aims (e.g., attempting to win an election or a war; see Bardall et al. (2020, 916)). As Bardall et al. (2020) write: "Attacks designed to disrupt the daily unfolding of political processes are different from the ever-present institutions, behaviors, and practices that enforce gender hierarchies (Sen et al., 2019). Norms that render women subordinate and "less-than," and the misogynistic desire to punish women who violate these norms (Manne, 2017), fuel the countless harms that befall women in public and private spaces, from dirty jokes to sexual assault" (918).

Literature on incivility has also been subject to much debate about whether it is a general phenomenon or a practice that may specifically target women. Most of the literature on incivility in politics has not taken gender or women as a dimension of focus. Classic articles like Mutz and Reeves (2005), Mutz (2007), and Brooks and Geer (2007) all explore the effects of incivility, particularly with regard to partisanship and polarization. Cortina (2008) takes a different approach: focusing on gendered incivility in the workplace, she departs from this "gender neutral" approach to studying incivility by arguing that incivility is "not 'general' at all but instead represents contemporary manifestations of gender and racial bias...one can mask discrimination (even without realizing it) behind everyday acts of incivility and still maintain an unbiased image" (55). Cortina posits that by applying ostensibly neutral uncivil behaviors—impatience, neglect, impoliteness, interruptions, condescending comments, and so on— selectively to women and racial minorities, especially those "who are perceived as highly competent and advancing in ways that threaten the dominant majority" (ibid, 65), men and white workers harboring explicit or implicit prejudices can ostracize, grind down, and eventually drive out those they want to exclude (see also Crenshaw 1991; Kuperberg 2018). Cortina et al. (2002) document that experiencing uncivil behaviors decreased job satisfaction, increased job stress, and increased willingness to exit practice among women attorneys (252).

Like VAWIP, digital misogyny, a newer field, defines its key term by the intended purpose of the behavior. Sobieraj (2020) argues that "digital misogyny. . . attempts to curtail women's freedom to use public spaces as equals" (4; see also Manne (2017)). The result is that many women opt out of public speech online, either altogether, or by minimizing their speech (selecting only certain fora, topics, etc. they believe are safer). Those that remain pay additional costs to ensure their safety: hiring firms to scrub their information from the internet, assistants to report threats and filter out hate mail, and avoiding posts with information about their plans or locations (including professional information like whether they will be giving a keynote address or meeting with readers).

APSR Submission Template

We argue that the most critical aspect of Krook (2020)'s theory of VAWIP for our paper—and the one that unites all three literatures, either explicitly or implicitly—is the decision to define these behaviors around their intended outcome (what Bardall et al. (2020) would call a "gendered motive," (919)), which is to exclude women from a given space. For VAWIP, that space is political (though what counts as a sufficiently political space has been contested, as we return to below). For scholars working on gendered incivility, those spaces are the workplace. For scholars of digital misogyny, those spaces are the Internet, and most commonly (though not exclusively) social media.

Research on these topics may in many cases seem empirically quite disparate. What, after all, do kidnappings of women city councilors in Mexico, tweets that American teen activist Emma Gonzalez is a "coconut head," and failures to invite women food workers out for drinks after work in Denmark have in common? Other than a shared target—women—the actions taken, the scope and severity of the behaviors, the fora, and the historical, political, economic, and cultural contexts vary.[3] What these acts share—and what even critics agree on—is that their intent is to subordinate and exclude women. Bardall et al. (2020) define this as a key contribution of the VAWIP approach: "the VAWIP approach shows how gender can motivate attacks...political violence can be gendered in multiple but distinct ways. Gendered motives appear when perpetrators use violence to preserve hegemonic men's control of the political system" (917).

This emphasis on subordination and exclusion as desired outcomes defines and distinguishes these literatures from scholarship that may seem ostensibly more related. For instance, all these disciplines—political science, sociology, psychology, and communication—have large literatures on sexism and gender stereotyping. What many of those literatures share, however, is a definition centered on the cognitive and social processes responsible for the attendant behaviors, rather than on the outcomes of those behaviors. At a theoretical level, that makes these literatures conceptually distinct. In turn, this is why we believe our manuscript should be situated within the VAWIP literature: our results speak more to outcomes than to specific cognitive processes.

Even without having data on voters' individual cognitive processes and beliefs, work by Sobieraj (2020) makes clear that we can distinguish the "gendered motives" (see Bardall et al. 2020, for discussion) that distinguish VAWIP from more general political violence through close reading of a given text:

> "Telling someone she is a filthy whore, for example, is intimate and ad hominem and yet decidedly generic...We may be able to distinguish the comments, for example, as the kind of abuse pointed at Black women, but be unable to distinguish what they have to do with *any particular* Black woman. This is a reminder; this abuse is structural, rooted in hostility toward the voice and visibility of individual speakers as *representatives* of specific groups of people" (5, emphases in the original).

Biroli (2018) concurs, writing: "Systemic violence targets women because they are women" (681). The "generic" quality that Sobieraj describes as a hallmark of structural abuse, i.e., abuse that treats individual women as representatives of all women, is clearly present in the texts the volunteers in our studies receive. When someone texts "fuck off Jessica you're a slut," or "no. your fat" or "bitch fuck u stop texting my fucking phone," having no more information about "Jessica" than her name, and despite "Jessica" using exactly the same text messages "Michael" and "Taylor" use to contact voters, we see clear evidence that the hostility is toward women as a group rather than some specific characteristic of this individual woman. They know nothing about this woman (who is of course not even necessarily a woman thanks to our randomized experiment). Despite and perhaps even because of knowing so little about Jessica, she provokes this reaction.

Looking beyond the behavior we observe in our studies, we theorize that Sobieraj's argument that women are representatives of their group, be it a specific subset of women or women in all their diversity, is also critical to the definition of VAWIP and merits further consideration by scholars in the field. Representation is deeply political (Pitkin 1967, Chs. 1, 4-6). Whether individual women "stand for" other women when they write #MeToo on Twitter, or "act for" other women when they take office—or simply "make [women] present" by sending political

[3]Equally, it is not obvious that all these behaviors should be classified as violence. Piscopo (2016), a critic of Krook's more expansive definition, writes "murdering female candidates and excluding female politicians from important meetings may not differ in kind—Krook and Restrepo Sanín (2016b) highlight how both acts seek the erasure of women from public life—but they differ in degree" (447). We return to the definition of violence in the next section.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

texts signed "Jessica" to other citizens—by Pitkin or Sobieraj's standards, these individual women may either see themselves as representing women, or be seen by others as representatives of women, and treated accordingly. To then focus VAWIP research exclusively on what Pitkin would call formalistic representation—candidacy, elections, votes, and so on—is to miss the diversity of political representation that women undertake and make, and the violence they may experience in doing so. We close this section by calling for a more expansive scope of research on VAWIP that examines the violence women experience as representatives regardless of the sphere in which that representation occurs.

Conceptualizing and Operationalizing Violence

As we are drawing upon three related but distinct literatures, it is important to clarify the concepts underlying the variables we employ. We explain below how our variables fit within and compare to existing concepts.

Within recent literature on VAWIP, the responses volunteers in our studies receive fall within the domains of psychological and semiotic violence (which as Krook (2020) notes, overlap). Psychological violence

> "seeks to disempower targets by degrading, demoralizing, or shaming them—often through efforts to instill fear, cause stress, or harm their credibility. Its varied forms comprise, but are not limited to, death threats, rape threats, intimidation, threats against family members, verbal abuse, bullying, rumor campaigns, illegal interrogation, surveillance, social ostracism, and blackmail. These acts may occur inside and outside official political settings and be carried out in person, by telephone, or via digital means like email and social media" (139).

Bardall (2013) advocates for a similar definition: "Psychological violence is an 'informal means of control [and] includes systematic ridicule, ostracism, shame, sarcasm, criticism, disapproval, exclusion and discrimination"' (Bardall 2010, cited in Bardall 2013, 2). Accordingly, semiotic violence may also constitute psychological violence (Krook 2020), but is distinguished by its effort to

> "mobiliz[e] semiotic resources—words, images, and even body language—to injure, discipline, and subjugate women...by communicating a message of group-based inferiority. Analyzed inductively, women's experiences two main modes of semiotic violence: *rendering women invisible*, attempting to 'symbolically annihilate' women in the public sphere, and *rendering women incompetent*, emphasizing 'role incongruity' between being a woman and being a leader" (187, emphases in the original).

Krook (2020) further identifies three subtypes of behaviors that attempt to render women invisible: "silencing," "not listening," and "manterrupting" (196-197). Silencing involves more formal or explicit modes of rendering women silent: for instance, not calling on women to speak, cutting them off, or sanctioning women for speaking. "Not listening" behavior is more informal, reducing women's voice by simply directing attention elsewhere. "Manterrupting" involves violating norms of alternating speakers when women are speaking and is also typically more informal behavior.

These strategies are notably similar to those Sobieraj (2020) identifies when studying digital misogyny: intimidating, shaming, and discrediting. We therefore focus on situating our variables within the categories of psychological and semiotic violence laid out by Krook.

Within these categories, the texts in our studies are most consistent with verbal abuse (psychological violence)—what we call our "offensiveness" measure—and "rendering women invisible" (semiotic violence)—what we call our "silencing" measure, which represents only texts that are both offensive and explicitly or implicitly request that contact should cease. Inoffensive or polite requests for contact to cease we call "withdrawal" to distinguish them from "silencing" (which is clearly uncivil); we discuss this decision further below.

Our "offensiveness" variable is labeled as such because we asked volunteers explicitly about how offensive the text was. It is also consistent with the language of the Equal Employment Opportunity Commission, which is in turned derived from Title VII of the Civil Rights Act of 1964:

APSR Submission Template

> "To be unlawful, the conduct must create a work environment that would be intimidating, hostile, or offensive to reasonable people. . . Offensive conduct may include, but is not limited to, offensive jokes, slurs, epithets or name calling, physical assaults or threats, intimidation, ridicule or mockery, insults or put-downs, offensive objects or pictures, and interference with work performance" (https://www.eeoc.gov/harassment).

Our "silencing" variable, which is derived from a function the organization uses to indicate that the respondent should not be contacted again, is somewhat more complicated. We attempt to position it here within Krook (2020)'s definition above. While we considered labeling our "silencing" variable "manterrupting," as it appears in some cases more consistent with this sort of informal behavior than e.g. formal sanctions, we felt both the text medium makes distinguishing interruption difficult (the norm is not so clearly to alternate between speakers), and many of those responding in this way are also women, making "manterrupting" seem a slightly misleading term.

We now turn to the literature on incivility because it has explored so many variables within what the VAWIP literature might call psychological or semiotic violence. Stryker et al. (2016) provides perhaps the most comprehensive approach to date, exploring 23 different measures of incivility: vulgarity, disrespect, slurs, interruption, name-calling, threatening harm, encouraging harm, misleading, preventing discussion, shouting, attacking character, attacking issue stance, failing to provide evidence in support of one's claims, exaggeration, making fun, attacking reputation, insulting, highlighting flubs, verbal jousting, rolling eyes, violating space, demonizing, and refusing to listen (542). Several of these are likely to fall within the purview of our offensiveness measure (vulgarity, disrespect, slurs, name-calling, shouting, insulting, demonizing, and verbal jousting). Two others are likely to map onto our silencing and withdrawal measures: preventing discussion and refusing to listen. Unfortunately, we do not have the time or resources to hand-code several thousand text responses into these 23 more fine-grained categories, but it would be fascinating to know whether there are gender differences on any of them. We encourage future research to explore this possibility further.

With regard to our decision to separate silencing and withdrawal, we again turn to the literature on incivility. Importantly, research on selective incivility emphasizes not just selective acts of aggression (what psychologists might term approach behaviors), but also of selective withdrawal: a "form of bias that Fiske (2002: 125) calls 'cool neglect,' or withholding 'basic liking and respect' rather than being openly hostile" Cortina (2008, 66). In our manuscript, we show evidence of gender differences in these two different types of behaviors: not just silencing (which Cortina might argue represents more overt bias or hostility), but also a sort of "cool neglect" through withdrawal.

This last point raises a final question for our paper, which is where to draw the limit on considering a behavior violence. Is polite withdrawal from an interaction violence against women in politics, for instance, even if the behavior only occurs when a woman is speaking? For our project, we draw the line between offensiveness or silencing (which are both consistent with standard definitions of violence in the VAWIP and digital misogyny scholarship), and withdrawals, which research on incivility argues may represent a form of gender discrimination, but do not, we think, rise to the level of violence. Beyond stating our stance as authors, disentangling violence from incivility on a theoretical level is beyond the scope of the present paper, especially given their many overlaps, questions of degree versus kind, and immense debates about the policy implications of doing so. We think exploring these questions would be a promising area for future research located at the intersection of these literatures.

Empirical Contribution

Studying violence against women in politics is difficult work.[4] Scholarship relying on archival work and interviews is constrained by how cost- and time-intensive it is to conduct such research. Given the extremely sensitive nature of such interviews, they cannot (for instance) be farmed out to local interviewers (see Krause 2020), making the costs that much higher. Accordingly, many qualitative researchers have focused on paradigmatic

[4]Though Ballington (2018) suggests it may be easier than studying violence against women in the home, which may be harder to systematically collect data on thanks to stigmas about reporting and privacy concerns (698).

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

cases that have a high payoff for understanding VAWIP, like the 2016 murder of British MP Jo Cox, the #MeToo movement among politicians, or the digital harassment of high-profile feminist activists.

Surveys, though in many cases equally expensive, take a wider lens and thus can study larger populations like municipal and regional parliaments in Sweden (Håkansson 2019), federal courts in the U.S. (Cortina et al. 2002), and New Zealand MPs (Every-Palmer et al. 2015). Surveys are nonetheless constrained by needing a list and contact information from which to generate a sampling frame, and it is far harder to find these for everyday activists and volunteers than it is to find them for sitting politicians (or even former candidates).

Nor are computational tools free from this issue: collecting hundreds of thousands of tweets using the Twitter API still requires pointing the API "at" something: a list of Canadian MP's handles (Rheault et al. 2019), say, or keywords like "bitch" (Blake et al. 2021), or hashtags like #SayHerName (Brown et al. 2017). In other words, to get misogynistic tweets about British MPs, you have to have a list of British MPs' names to collect anything (Ward and McLoughlin 2020). Attempts to study these topics in the general population—e.g., to study digital sexism between citizens writ large—run into other types of problems. For instance, Blake et al. (2021) collect 1.8 billion randomly sampled tweets to study digital misogyny, and according to their coding scheme, identify only 16,791 that meet their criteria for misogyny (.0009 percent). Small wonder that most researchers hoping to understand VAWIP, even quantitatively, focus on those most likely to experience threats, harassment, and violence. Håkansson (2019), for instance, finds that nearly 70% of women mayors in Sweden experience some form of violence every year. Yet Håkansson (2019) also theorizes and subsequently demonstrates that lower visibility (e.g., rank-and-file politicians) substantially reduces the amount of violence women experience (40). Thomas et al. (2019) theorize and find similarly that increased visibility results in increased violence among American mayors.

Our research therefore complements existing empirical scholarship by allowing us to examine VAWIP in a context and sample that is hard for many of these methodological tools to observe Ballington (see 2018, 700): everyday activism through political volunteering. While VAWIP researchers have theorized for some time that less elite women working in politics experience violence, it is costly to document. We believe our findings provide empirical support for including women engaged in less formalized types of political work within the broader topic of VAWIP. We pursue the implications of our empirical findings in the section that follows.

Implications for the Future Study of VAWIP

Should VAWIP scholars treat violence against activists, journalists, or even participants in everyday political deliberation as VAWIP? As we argued above, we believe there is theoretical reason to consider women to be political representatives even outside of formal political roles and institutions. We therefore join Alam (2021), Biroli (2018), and Krook (2020) in calling for an expansive definition:

> "The concept of violence against women in politics...has largely been restricted to actions perpetrated against women in elections and/or within formal political institutions...however, parallel campaigns have surfaced to draw attention to violence committed against women human rights defenders and against female journalists, respectively. These efforts take up highly similar issues concerning violence as a barrier to women's participation in the political field. This book advocates for joining these various streams to forge a more comprehensive concept of violence against women in politics" (ibid, 36).

Following Krook (2020), Håkansson (2019), and Thomas et al. (2019), we show that violence does indeed exist along a spectrum of visibility in politics: the more visible, the more violence.[5] Our "meso-level" volunteers experience less violence than politicians, candidates, and high-profile activists, but more violence than Blake et al. (2021) seem to document among the entire population of Twitter users.

Moreover, Sobieraj (2020) notes that women activists experiencing what Krook and Restrepo Sanín (2016a) would call psychological and semiotic violence on social media are told to "just get off the Internet" (Sobieraj

[5]Strid et al. (2013) theorize that visibility shapes our perceptions of representation by the multiply marginalized, who may simultaneously be more visible and more likely to be overlooked. We think this idea is interesting and merits expansion in future research.

2020, 20). Our research shows that the Internet is not the problem. Speaking out on political issues even in a comparatively safe space—texting requests to other members of one's own political organization—still unleashes more hostility towards women than men. Our standardized messages and random assignment of names to volunteers make clear that the issue is not how women "do" politics, nor their selection of a venue in which to participate politically, nor is it that violence is somehow the inevitable cost of political candidacy or elite activism. The issue is resistance to women's voices.

Finally, in addition to our call above to theorize and study how VAWIP manifests beyond the formal sphere of electoral politics and policymaking, we argue that broadening the definition of VAWIP will help break new ground by enabling the study of the effects of violence at all stages of what scholars of gendered political ambition might call the "pipeline" to office (Thomsen and King 2020; Bernhard et al. 2021). This is because the intent of VAWIP is to exclude women from participating in the political sphere. If scholars of VAWIP do not study what happens to women before they formally enter politics (e.g., by filing for candidacy)—and Sobieraj (2020) documents that many women activists reduce or end their advocacy altogether after experiencing sustained misogyny and threats—we may miss important ways in which violence against women shapes who enters the formal political arena.

Supplemental References

Alam, Zainab. 2021. Violence against women in politics: The case of Pakistani women's activism. *Journal of Language Aggression and Conflict*, 21–46.

Anderson, Ashley A., Sara K. Yeo, Dominique Brossard, Dietram A. Scheufele, and Michael A. Xenos. 2016. Toxic Talk: How Online Incivility Can Undermine Perceptions of Media. *International Journal of Public Opinion Research*, edw022.

Araújo, Victor and Malu AC Gatto. 2022. Can conservatism make women more vulnerable to violence? *Comparative Political Studies*, 00104140211024313.

Ballington, Julie. 2018. Turning the Tide on Violence against Women in Politics: How Are We Measuring Up? *Politics & Gender 14*(4), 695–701.

Bardall, Gabrielle. 2013. Gender-Specific Election Violence: The Role of Information and Communication Technologies. *Stability: International Journal of Security & Development 2*(3), 60.

Bardall, Gabrielle and Gabrielle Bardall. 2020. Symbolic Violence as a Form of Violence against Women in Politics: A Critical Examination. *Revista mexicana de ciencias políticas y sociales 65*(238), 379–389.

Bardall, Gabrielle, Elin Bjarnegård, and Jennifer M Piscopo. 2020. How is Political Violence Gendered? Disentangling Motives, Forms, and Impacts. *Political Studies 68*(4), 916–935.

Beer, Caroline. 2017. Left Parties and Violence against Women Legislation in Mexico. *Social Politics 24*(4), 511–537.

Bernhard, Rachel, Shauna Shames, and Dawn Langan Teele. 2021. To Emerge? Breadwinning, Motherhood, and Women's Decisions to Run for Office. *American Political Science Review 115*(2), 379–394.

Berthet, Valentine and Johanna Kantola. 2021. Gender, Violence, and Political Institutions: Struggles over Sexual Harassment in the European Parliament. *Social Politics: International Studies in Gender, State & Society 28*(1), 143–167.

Biroli, Flávia. 2018. Violence against Women and Reactions to Gender Equality in Politics. *Politics & Gender 14*(4), 681–685.

Bjarnegård, Elin. 2021. The continuum of election violence: Gendered candidate experiences in the Maldives. *International Political Science Review*, 019251212097711.

Bjarnegård, Elin, Sandra Håkansson, and Pär Zetterberg. 2020. Gender and Violence against Political Candidates: Lessons from Sri Lanka. *Politics & Gender*, 1–29.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

Blake, Khandis R., Siobhan M. O'Dean, James Lian, and Thomas F. Denson. 2021. Misogynistic Tweets Correlate With Violence Against Women. *Psychological Science 32*(3), 315–325.

Brooks, Deborah Jordan and John G. Geer. 2007. Beyond Negativity: The Effects of Incivility on the Electorate. *American Journal of Political Science 51*(1), 1–16.

Brown, Melissa, Rashawn Ray, Ed Summers, and Neil Fraistat. 2017. #SayHerName: A case study of intersectional social media activism. *Ethnic and Racial Studies 40*(11), 1831–1846.

Brown, Nadia, Rebecca Gill, Stella Rouse, and Elizabeth Sharrow. 2019. NSF Award Search: Award # 1936086 - ADVANCE Partnership: #MeTooPoliSci Leveraging A Professional Association to Address Sexual Harassment in Political Science. `https://www.nsf.gov/awardsearch/showAward?AWD_ID=1936086&HistoricalAwards=false`. (Accessed on 05/19/2021).

Brulé, Rachel E.. 2020. *Women, Power, and Property: The Paradox of Gender Equality Laws in India*. Cambridge University Press.

Center, Pew Research. 2017. Online Harassment 2017.

Cickaric, Lilijana and Zorica Mrsevic. 2019. *Contemporary issues and perspectives on gender research*.

Collier, Cheryl N and Tracey Raney. 2018. Understanding Sexism and Sexual Harassment in Politics: A Comparison of Westminster Parliaments in Australia, the United Kingdom, and Canada. *Social Politics: International Studies in Gender, State & Society 25*(3), 432–455.

Collignon, Sofia and Wolfgang Rüdig. 2020. Harassment and Intimidation of Parliamentary Candidates in the United Kingdom. *The Political Quarterly 91*(2), 422–429.

Collignon, Sofia and Wolfgang Rüdig. 2021. Increasing the cost of female representation? The gendered effects of harassment, abuse and intimidation towards Parliamentary candidates in the UK. *Journal of Elections, Public Opinion and Parties 31*(4), 429–449.

Córdova, France. 2020. Leadership to change a culture of sexual harassment. *Science 367*(6485), 1430–1431.

Cortina, Lilia M.. 2008. Unseen Injustice: Incivility as Modern Discrimination in Organizations. *The Academy of Management Review 33*(1), 55–75.

Cortina, Lilia M., Kimberly A. Lonsway, Vicki J. Magley, Leslie V. Freeman, Linda L. Collinsworth, Mary Hunter, and Louise F. Fitzgerald. 2002. What's Gender Got to Do with It? Incivility in the Federal Courts. *Law & Social Inquiry 27*(2), 235–270.

Cortina, Lilia M., Vicki J. Magley, Jill Hunter Williams, and Regina Day Langhout. 2001. Cortina: Incivility in the workplace: Incidence and impact. - Google Scholar. *Journal of Occupational Health Psychology 6*(1), 64–80.

Crenshaw, Kimberle. 1991. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review 43*(6), 1241.

Dalton, Emma. 2017. Sexual harassment of women politicians in Japan. *Journal of Gender-Based Violence 1*(2), 205–219.

Druckman, James N., S. R. Gubitz, Matthew S. Levendusky, and Ashley M. Lloyd. 2019. How Incivility on Partisan Media (De)Polarizes the Electorate. *Journal of Politics 81*(1), 291–295.

Erikson, Josefina, Sandra Håkansson, and Cecilia Josefsson. 2021. Three Dimensions of Gendered Online Abuse: Analyzing Swedish MPs' Experiences of Social Media. *Perspectives on Politics*, 1–17.

Erikson, Josefina and Cecilia Josefsson. 2019. The legislature as a gendered workplace: Exploring members of parliament's experiences of working in the Swedish parliament. *International Political Science Review 40*(2), 197–214.

Esposito, Eleonora. 2021. Introduction: Critical perspectives on gender, politics and violence. *Journal of Language Aggression and Conflict 9*(1), 1–20.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

Esposito, Eleonora and Sole Alba Zollo. 2021. "How dare you call her a pig, I know several pigs who would be upset if they knew"*: A multimodal critical discursive approach to online misogyny against UK MPs on YouTube. *Journal of Language Aggression and Conflict 9*(1), 47–75.

Every-Palmer, Susanna, Justin Barry-Walsh, and Michele Pathé. 2015. Harassment, stalking, threats and attacks targeting New Zealand politicians: A mental health issue. *Australian & New Zealand Journal of Psychiatry 49*(7), 634–641.

Folke, Olle, Johanna Rickne, Seiki Tanaka, and Yasuka Tateishi. 2020. Sexual Harassment of Women Leaders. *Daedalus 149*(1), 180–197.

Gervais, Bryan T.. 2015. Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-based Experiment. *Journal of Information Technology & Politics 12*(2), 167–185.

Harmer, Emily and Rosalynd Southern. 2021. Digital microaggressions and everyday othering: an analysis of tweets sent to women members of Parliament in the UK. *Information, Communication & Society 24*(14), 1998–2015.

Herrick, Rebekah and Sue Thomas. 2021a. Not Just Sticks and Stones: Psychological Abuse and Physical Violence among U.S. State Senators. *Politics & Gender*, 1–26.

Herrick, Rebekah and Sue Thomas. 2021b. Violence among state house candidates during the covid-19 pandemic. *Legislative Studies Quarterly*.

Herrick, Rebekah, Sue Thomas, Lori Franklin, Marcia L. Godwin, Eveline Gnabasik, and Jean Reith Schroedel. 2021. Physical violence and psychological abuse against female and male mayors in the United States. *Politics, Groups, and Identities 9*(4), 681–698.

Hershcovis, M. Sandy. 2011. "Incivility, social undermining, bullying. . . oh my!": A call to reconcile constructs within workplace aggression research. *Journal of Organizational Behavior 32*(3), 499–519.

Holland, Kathryn J. and Lilia M. Cortina. 2013. When Sexism and Feminism Collide: The Sexual Harassment of Feminist Working Women. *Psychology of Women Quarterly 37*(2), 192–208.

Håkansson, Sandra. 2019. Do Women Pay a Higher Price for Power? Gender Bias in Political Violence in Sweden. *Journal of Politics*.

Johnson-Myers, Tracy-Ann. 2021. Violence against women in politics: Female politicians' experience with political violence in Jamaica. *Bulletin of Latin American Research*.

Kabat-Farr, Dana and Lilia M. Cortina. 2012. Selective Incivility: Gender, Race, and the Discriminatory Workplace. In *Gender and the Dysfunctional Workplace*. Edward Elgar Publishing.

Karpowitz, Christopher F and Tali Mendelberg. 2014. *The Silent Sex: Gender, Deliberation, and Institutions*. Princeton: Princeton University Press.

Karpowitz, Christopher F, Tali Mendelberg, and Lee Shaker. 2012. Gender inequality in deliberative participation. *American Political Science Review 106*(3), 533–547.

Kenski, Kate, Kevin Coe, and Stephen A. Rains. 2017. Perceptions of Uncivil Discourse Online: An Examination of Types and Predictors. *Communication Research*.

Kim, Jin Woo, Andrew Guess, Brendan Nyhan, and Jason Reifler. 2021. The Distorting Prism of Social Media: How Self-Selection and Exposure to Incivility Fuel Online Comment Toxicity. *Journal of Communication*, 25.

Krause, Jana. 2020. Restrained or constrained? Elections, communal conflicts, and variation in sexual violence. *Journal of Peace Research 57*(1), 185–198.

Krook, Mona Lena. 2017. Violence Against Women in Politics. *Journal of Democracy 28*(1), 74–88.

Krook, Mona Lena. 2018. Violence against Women in Politics: A Rising Global Trend. *Politics & Gender 14*(4), 673–675.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

Krook, Mona Lena. 2019a. Global feminist collaborations and the concept of violence against women in politics. *Journal of International Affairs 72*(2), 77–94.

Krook, Mona Lena. 2019b. New Horizons in Women's Political Rights. In S. Franceschet, M. L. Krook, and N. Tan (Eds.), *The Palgrave Handbook of Women's Political Rights*, pp. 73–81. London: Palgrave Macmillan UK.

Krook, Mona Lena. 2020a. *Violence against Women in Politics*. Oxford: Oxford University Press.

Krook, Mona Lena. 2020b. Violence against women in politics. In *How Gender Can Transform the Social Sciences*, pp. 57–64. Springer.

Krook, Mona Lena. 2022. Semiotic Violence against Women: Theorizing Harms against Female Politicians. *Signs: Journal of Women in Culture and Society 47*(2), 371–397.

Krook, Mona Lena and Juliana Restrepo Sanín. 2016a. Gender and political violence in Latin America. Concepts, debates and solutions. *Política y gobierno 23*(1), 127–162.

Krook, Mona Lena and Juliana Restrepo Sanín. 2016b. Violence Against Women in Politics. A Defense of the Concept. *Política y gobierno 23*(2), 459–490.

Krook, Mona Lena and Juliana Restrepo Sanín. 2020. The cost of doing politics? Analyzing violence and harassment against female politicians. *Perspectives on Politics 18*(3), 740–755.

Kuperberg, Rebecca. 2018. Intersectional Violence against Women in Politics. *Politics & Gender 14*(4), 685–690.

Manne, Kate. 2017. *Down girl: The logic of misogyny*. Oxford: Oxford University Press.

Mayaram, Shail. 2002. New modes of violence: The backlash against women in the Panchayat system. *The Violence of Development: The Politics of Identity, Gender and Social Inequalities in India. New Delhi: Kali for Women 393*, 424.

Mechkova, Valeriya and Steven L. Wilson. 2021. Norms and rage: Gender and social media in the 2018 U.S. mid-term elections. *Electoral Studies 69*, 102268.

Meltzer, Kimberly. 2015. Journalistic Concern about Uncivil Political Talk in Digital News Media: Responsibility, Credibility, and Academic Influence. *The International Journal of Press/Politics 20*(1), 85–107.

Miner, Kathi N. and Angela Eischeid. 2012. Observing Incivility toward Coworkers and Negative Emotions: Do Gender of the Target and Observer Matter? *Sex Roles 66*(7), 492–505.

Muddiman, Ashley. 2017. Personal and public levels of political incivility. *International Journal of Communication 11*, 21.

Mutz, Diana C.. 2007. Effects of "In-Your-Face" Television Discourse on Perceptions of a Legitimate Opposition. *American Political Science Review 101*(4), 621–635.

Mutz, Diana C. and Byron Reeves. 2005. The New Videomalaise: Effects of Televised Incivility on Political Trust. *American Political Science Review 99*(1), 1–15.

Och, Malliga. 2020. Manterrupting in the German Bundestag: Gendered Opposition to Female Members of Parliament? *Politics & Gender 16*(2), 388–408.

Okimoto, Tyler G. and Victoria L. Brescoll. 2010. The Price of Power: Power Seeking and Backlash Against Female Politicians. *Personality and Social Psychology Bulletin 36*(7), 923–936.

Papacharissi, Zizi. 2004. Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society 6*(2), 259–283.

Pearson, Christine M, Lynne M Andersson, and Christine L Porath. 2000. Assessing and attacking workplace incivility. *Organizational Dynamics 29*(2), 123–137.

Piquero, Nicole Leeper, Alex R. Piquero, Jessica M. Craig, and Stephen J. Clipper. 2013. Assessing research on workplace violence, 2000–2012. *Aggression and Violent Behavior 18*(3), 383–394.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

Piscopo, Jennifer M. 2016. Capacidad estatal, justicia criminal y derechos políticos: nueva mirada al debate sobre la violencia contra las mujeres en política. *Política y gobierno 23*(2), 437–458.

Prillaman, Soledad Artiz. 2024. *Agency, Participation, and Power: Breaking the Patriarchal Household through Collective Action*. Cambridge: Cambridge University Press.

Restrepo Sanín, Juliana. 2018. The Law and Violence against Women in Politics. *Politics & Gender 14*(4), 676–680.

Restrepo Sanín, Juliana. 2020. Criminalizing Violence against Women in Politics: Innovation, Diffusion, and Transformation. *Politics & Gender*, 1–32.

Rheault, Ludovic, Erica Rayment, and Andreea Musulan. 2019. Politicians in the line of fire: Incivility and the treatment of women on social media. *Research & Politics 6*(1), 1–7.

Rowe, Ian. 2015. Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society 18*(2), 121–138.

Safir, Alexandra Z and Mayesha Alam. The 2015 Nigeria elections and violence against women in politics. *Georgetown Institute for Women, Peace, and Security*, 1–5.

Samnani, Al-Karim and Parbudyal Singh. 2012. 20 Years of workplace bullying research: A review of the antecedents and consequences of bullying in the workplace. *Aggression and Violent Behavior 17*(6), 581–589.

Sanín, Juliana Restrepo. 2020. Violence against Women in Politics: Latin America in an Era of Backlash. *Signs: Journal of Women in Culture and Society 45*(2), 302–310.

Schneider, Paige and David Carroll. 2020. Conceptualizing more inclusive elections: violence against women in elections and gendered electoral violence. *Policy Studies 41*(2-3), 172–189.

Settles, Isis H. and Rachel C. O'Connor. 2014. Incivility at Academic Conferences: Gender Differences and the Mediating Role of Climate. *Sex Roles 71*(1-2), 71–82.

Sobieraj, Sarah. 2018. Bitch, slut, skank, cunt: patterned resistance to women's visibility in digital publics. *Information, Communication & Society 21*(11), 1700–1714.

Sobieraj, Sarah. 2020. *Credible threat: Attacks against women online and the future of democracy*. Oxford: Oxford University Press.

Sobieraj, Sarah and Jeffrey M. Berry. 2011. From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and Cable News. *Political Communication 28*(1), 19–41.

Strid, S., S. Walby, and J. Armstrong. 2013. Intersectionality and Multiple Inequalities: Visibility in British Policy on Violence Against Women. *Social Politics: International Studies in Gender, State & Society 20*(4), 558–581.

Stryker, Robin, Bethany Anne Conway, and J. Taylor Danielson. 2016. What is political incivility? *Communication Monographs 83*(4), 535–556.

Sydnor, Emily. 2018. Platforms for Incivility: Examining Perceptions Across Different Media Formats. *Political Communication 35*(1), 97–116.

Theocharis, Yannis, Pablo Barberá, Zoltán Fazekas, and Sebastian Adrian Popa. 2020. The Dynamics of Political Incivility on Twitter. *SAGE Open 10*(2), 2158244020919447.

Thomas, Sue, Rebekah Herrick, Lori D. Franklin, Marcia L. Godwin, Eveline Gnabasik, and Jean R. Schroedel. 2019. Not for the Faint of Heart: Assessing Physical Violence and Psychological Abuse against U.S. Mayors. *State and Local Government Review 51*(1), 57–67.

Thomsen, Danielle M and Aaron S King. 2020. Women's representation and the gendered pipeline to power. *American Political Science Review 114*(4), 989–1000.

Ward, Stephen and Liam McLoughlin. 2020. Turds, traitors and tossers: The abuse of UK MPs via Twitter. *Journal of Legislative Studies 26*(1), 47–73.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

DETAILED METHODS AND MATERIALS

Sampling Frame

The sample for the first experiment is 60,356 individuals who had interacted with NextGen America in the past; for the second experiment, 75,231 individuals drawn from the same population. The organization thought these individuals were more likely to attend a given political event.

Treatments

After creating the sample, individual supporters were randomly assigned to one of four conditions. Individuals in the "male-name" condition received a SMS message from a volunteer assigned to use the name Michael. Individuals in the "female-name" condition received a message from a volunteer assigned to use the name Jessica. Those in the "ambiguous-gender" condition received a message from a volunteer assigned to use the name Taylor. Individuals in the "unnamed" condition received messages with no name given.

Volunteers did not know about the experiment's purpose. They were simply informed that the organization was trying something new to improve their messaging program. After the studies were completed, all staff and volunteers were debriefed, and the organization received a copy of our report.

After randomizing individual supporters to one of the four conditions, batches of texts were given to volunteers to send. For example, a volunteer could be given 1,000 messages in the female-name condition. If they finish all 1,000 texts, they might be given another batch of messages in the same condition, or they may be given a batch of messages in a different condition. If a volunteer fails to finish their assigned texts, then NGA staffers reclaim the unfinished texts and reassign them to another volunteer. This process ensures the relationship between the assignment of names and individuals is random, but there can be imbalances in how many texts in each condition are successfully sent by a given volunteer. Models subsetting to supporters who were sent messages or supporters who responded to texts will suffer from post-treatment bias.

These names signal their intended genders. Based on Census data, "Jessica" has a 99.7% probability of being a woman, "Michael" has a 99.5% chance of being a man, and "Taylor" has a 75.4% probability of being a woman. We also conducted our own study. We recruited a convenience sample of 260 individuals through Amazon's Mechanical Turk platform and asked them, "if you had to guess, would you guess [NAME] was a woman, a man, or unsure?" We randomized the order the names were presented in. One hundred percent of the 260 respondents guessed that "Jessica" was a woman and 98.5% guessed that "Michael" was a man. The most common response (49.2%) for "Taylor" was "unsure," and the remainder of the responses were divided between woman (41%) and man (9.8%). We thus remain confident that Taylor is a more ambiguously gendered name than "Jessica" or "Michael."

Messages

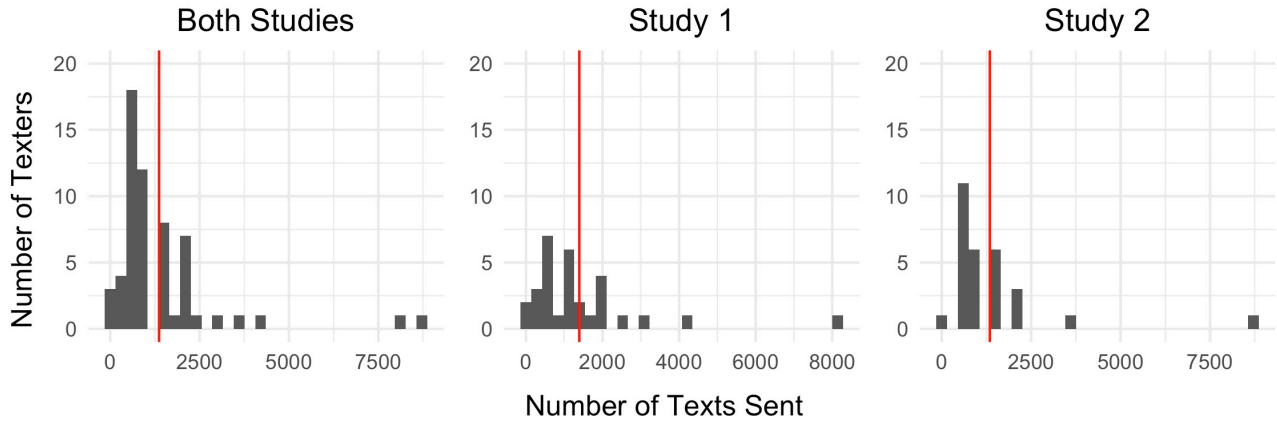
Study 1 Hi [VOTER'S NAME], I'm [EXPERIMENTAL TREATMENT NAME], a volunteer with NextGen America. This Saturday, March 24, young people have organized a march to spread awareness and demand gun safety legislation from our elected officials. Find a local March For Our Lives here: http://nxtgn.us/dkh - Can we count on you to join us at the march?

Study 2 Hey [VOTER'S NAME], it's [EXPERIMENTAL TREATMENT NAME] volunteering w/ NextGen America. EPA Administrator Scott Pruitt is under investigation for his waste of taxpayer dollars and shady ties to corporate polluters. Urge your members of Congress to call on Scott Pruitt to resign: http://NXTGN.US/du5

Number of Texts Sent Figure S1 shows how many messages volunteers sent. The average volunteer sent approximately 1,300 messages, which means that we can measure quite precisely how the same volunteer is treated when using each name (or no name).

APSR Submission Template

FIGURE S1. Number of texts sent. Figure shows how many texts were sent by volunteer across both studies and for each study. The red line indicates the average number of texts sent.



Dependent Variables

We measure our first dependent variable, *offensiveness*, in two ways. The measure in the first study is a self-report: how offensive volunteers felt the responses they received from each individual voter were. In the messaging system, after volunteers correspond with a voter, volunteers are asked whether the response they received was "offensive" or not. If the volunteer felt the message was offensive, then they marked "Yes." If the volunteer marked "Yes," then the respondent was coded as a 1, and if not, then they were coded as a 0. The measure of "offensiveness" in the second study is generated by ratings from two independent women coders. Each coder rated the messages on a five-point scale from "non-offensive" (1), to "very offensive" (5). The inter-coder Pearson's correlation was .74, which suggests moderate agreement on what constitutes offensiveness between the two volunteers. To compare scores between the two studies (main results reported in the paper), we also created a binary measure for Study 2: a message would be marked as offensive if either coder had marked it as offensive (100), and inoffensive if neither had (0). Our offensiveness findings for Study 2 hold whether the five-point scale is used or the binary scale is used, so for brevity we report only the binary analyses in the main paper, which allows us to condense the findings for both studies.

Our second original dependent variable, *discouragement*, collected only in the second study, measures whether a message encourages or discourages a volunteer from volunteering again in the future. Two independent women coders rated each message on a scale from "really encouraging" (1), to "really discouraging" (7). The Pearson's correlation between the two coders was .95, which suggests a high level of agreement between the two volunteers. We again summed and rescaled from 0 to 100 the two coders' ratings. To ensure our estimates of discouraging behavior are conservative, when a voter did not respond, we coded them as a 1 (really encouraging).

Our third dependent variable, *silencing*, measures whether a voter uses offensive or intimidating language to get the volunteer to cease contact. Organization rules instruct volunteers to "opt-out" respondents from future contact if the respondent requests the volunteer to stop contacting them, or if the respondent harasses the volunteer. If the respondent was marked in the system as "opted-out" and had an offensive score greater than 0, they were given a 100, and if not, a 0. We treat this as a measure for "silencing" behavior because it prevents volunteers from any future outreach to the individual and is meant to intimidate the volunteer. Because this variable is a subset of offensiveness, we should expect effects for silencing to be weaker than effects for offensiveness.

Our fourth dependent variable, *withdrawal*, measures whether a voter ends all future activist outreach but does not use offensive language to do so. If so, they were scored as a 100 on withdrawal, and if not, a 0.

We report statistics for Pearson's r and Krippendorff's alpha to evaluate inter-coder reliability. The correlation between the two coders for the offensiveness scale was .74; on the discouragement scale, .95. The Krippendorff's alpha for the offensiveness scale was 0.73; for the discouragement scale, 0.49.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

We also evaluate whether offensiveness, silencing, discouragement, and withdrawal measure different concepts. In Study 1, the correlation between volunteer-coded offensiveness and silencing was 0.77. In Study 2, the correlation between coder-rated offensiveness and silencing was 0.66. This reflects our conceptualization of silencing as an offensive request by the respondent to end all future activism from that volunteer. The correlation between volunteer-coded offensiveness and withdrawal was -0.01 in Study 1. The correlation between volunteer-coded offensiveness and withdrawal was -0.003 in Study 2. These low correlations reflect our conceptualization of withdrawal as polite or non-offensive requests to end all future activism. The correlation between coder-rated offensiveness and coder-rated discouragement was 0.51 in Study 2. This suggests meaningful differences between the concepts. The correlation between silencing and discouragement was 0.47 in Study 2, suggesting they are distinct concepts. The correlation between withdrawal and discouragement was 0.61 in Study 2. This is consistent with what we expect because requests to end all future outreach are likely disheartening even if they are not offensive. Finally, the overall correlation between silencing and withdrawal was 0.93. The high correlation reflects that silencing is a combination of both ending outreach and offensiveness.

Coding Instructions Coders saw the following instructions:

"You will be given individual response text messages from people who originally received text messages from NGA. You will undertake two coding tasks for each text:

Task 1: How offensive does each text seem? Each message is to be coded on a 1 to 5 scale where 1 is non-offensive and 5 is very offensive. Below illustrates what the scale looks like:

1. Non-offensive
2. Slightly offensive
3. Moderately offensive
4. Fairly offensive
5. Very offensive

Task 2: Would this message encourage or discourage you from volunteering in the future? Each message is to be coded on a 1 to 7 scale where 1 is really encourage and 7 is really discourage. Below illustrates what the scale looks like:

1. Really encourage
2. Somewhat encourage
3. Slightly encourage
4. Neither encourage nor discourage
5. Slightly discourage
6. Somewhat discourage
7. Really discourage

Each of you will be given the same list of text messages. However, you should not consult each other while coding because this could bias the results. Your codings don't have to match each other's because we will make an average "rating" from your codings, so it's not an error if you disagree.

If either of you have a question or comment, let me know and I'll help you on the individual text message. Some text messages may be broken up because of processing errors on Relay's side; if you find this, let me know and I'll help you."

Alternative Approaches to Measurement In this section, we report information on alternative measurement approaches we used to assess the reliability of our indicators.

Human-Coded Approaches

We attempted to replicate both our measures of offensiveness: in Study 1, dichotomized offensiveness, and in Study 2, the five-point scale (dichotomized for presentation in the main paper). Our concern was that, by having

APSR Submission Template

two different measurement scales, it would be hard for careful readers to assess whether they were comparing "apples to apples" in both studies, even with the consistent results.

We also attempted to replicate our seven-point scale measure of discouragement, which we originally only collected for Study 2. Here, our concern was that we might find something quite different in Study 1, limiting the generalizability of our findings. We tried replicating both our offensiveness measures and seven-point discouragement measure using two new sets of coders: first, online survey respondents recruited through Prolific, a sampling firm (Study A); and second, three research assistants with previous experience as political volunteers (Study B). Both sets were recruited in summer 2020. All texts from Study 1 and Study 2 were included, but to respect our non-disclosure agreement, all identifying information was stripped from the texts before they were shared. This means, for instance, a text saying "This isn't John's phone" became "This isn't [NAME]'s phone."

In Study A, 708 survey respondents recruited through Prolific were randomly assigned to see 10 of the response text messages from the original studies, each on its own page. Each page first asked them to rate the offensiveness of the text on the five-point scale above ("How offensive does the text seem?"). Below, they were asked to rate how discouraging the text was on the seven-point scale described above, but with a slightly differently worded question: "If you were a volunteer for a political campaign, would this message encourage or discourage you from volunteering in the future?" We modified the text of the question in order to clarify the intent of the encouragement/discouragement question for individuals who might have no political volunteer experience or lack necessary context for the question. We then averaged the results from the coders to create the dependent variables. Due to randomization, the average text message was coded by two individuals, but some texts were not coded, and others were coded by more than two individuals.

In Study B, three undergraduate research assistants recruited based on their prior political volunteer experience were asked to code the binary offensiveness measure ("How offensive does the text seem?") from Study 1 and the seven-point discouragement measure ("Would this text encourage or discourage you from volunteering in the future?") from Study 2. Note that the discouragement question wording here is the original from Study 2, not the slightly modified version used for the Prolific respondents (who may not have had political volunteer experience). All three research assistants coded all received response text messages. As before, for each text we average the codings for each measure across the research assistants. The three research assistants were women.

We provide the results for both studies below. Broadly, we find that our results for more offensive treatment of female-named volunteers replicate in both Study A and Study B. We therefore feel confident in the original findings for offensiveness is appropriate.

However, our discouragement measure did not replicate in Study A or Study B, even when restricted just to texts from Study 2 (where we originally found significant discouragement). As we explain below, while we cannot rule out the possibility that this is some sort of issue with the clarity of the instructions for our coders in Study A or Study B, we take the conservative approach of providing all the data on discouragement in these appendices, but refrain from making any claims about discouragement in the main paper.

Table S1 reports the Pearson correlation between intern-coded and volunteer-coded offensive and discouraging message ratings with the ratings provided by Prolific raters and research assistants. We subset to only ratings when respondents texted the volunteer back. This gives us a more conservative correlation because we code non-responses as 0 (inoffensive) for offensiveness and 50 (neutral) for discouragement. Column 1 shows that the Prolific offensive message ratings were moderately correlated with our original volunteer-coded and intern-coded offensive message ratings. The correlation improves when research assistants rate offensive messages, but the correlation is not strong. Column 2 displays the same correlations for discouraging message ratings only for Study 2 because we did not collect discouraging message ratings for the first experiment. Prolific discouraging message ratings were weakly related to our intern-coded messages. The research assistant discouraging message ratings were much more strongly correlated with the intern-coded discouraging message ratings than the Prolific ratings, but they were not incredibly strong correlations.

Table S1 also shows the Pearson correlations between the original NGA volunteer-coded discouragement measures and the measure when coded by Prolific respondents and our research assistants. The correlation between the Prolific-coded measures was very poor at 0.19. The correlation between our research assistant-coded measures and the original measure was 0.30. Taken together, these correlations were not strong enough for us to

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

TABLE S1. Correlations between Original and New Ratings

Sample	Offensive Messages (Both Studies)	Discouraging Messages (Study 2)
Online Survey Respondents	0.45	0.19
Research Assistants	0.64	0.30

have confidence in our discouragement measure.

Robustness of Offensiveness Codings

Study 2 - Ordinal measure of offensiveness Table S2 shows the regression results for Study 2, which used independent raters' codings of offensiveness. In the first column, we show that female-named and unnamed volunteers were significantly more likely to receive offensive replies than an ambiguously-named volunteer. Male-named volunteers were slightly less likely to receive offensive replies than the ambiguously-named volunteer. Looking at columns two and three, we see that female-named volunteers fare slightly worse among women voters than they do among men voters, but the estimate remains substantively unchanged. Male-named volunteers fare equally well with men and women voters. Unnamed volunteers are treated quite differently by men and women: they are much more likely to receive offensive content from men than from women.

TABLE S2. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving an Offensive Message, as Coded by Other Volunteers

	Offensiveness		
	Full Sample	Women only	Men only
Female Name	0.040^{*}	0.052^{*}	0.034
	(0.020)	(0.024)	(0.037)
Male Name	−0.022	−0.023	−0.027
	(0.018)	(0.017)	(0.036)
No Name	0.094^{***}	0.050^{*}	0.148^{**}
	(0.026)	(0.024)	(0.053)
Constant	0.062^{***}	0.040^{**}	0.093^{***}
	(0.012)	(0.013)	(0.024)
N	75,231	41,143	29,321
Adjusted R^2	0.0003	0.0003	0.0005

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Robust standard errors are reported in parentheses.

Different ways of dichotomizing offensiveness Below, we investigate whether the way we dichotomized the offensive message rating is driving our results. Table S3 presents the results from analyzing dichotomized offensive replies following different coding rules (in Study 2, we originally measure offensiveness using a 5-pt scale, taking the average of the two coders' ratings). We vary whether we code a message as offensive when the offensive index measure is greater than 50, 40, 30, and 20. We show that our results are broadly consistent with our main results and are not sensitive to our coding decisions. Across all four columns, female-named and unnamed volunteers were significantly more likely to receive an offensive messages compared to the ambiguously-named volunteer. Across all four columns, male-named volunteers were less likely to receive an offensive message compared to ambiguously-named volunteers, but none of these effects were statistically significant.

APSR Submission Template

TABLE S3. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving an Offensive Message, Both Studies - Robustness Check

	Offensive Replies			
	Rated > 50	> 40	> 30	> 20
Female Name	0.076**	0.076**	0.076*	0.076*
	(0.029)	(0.029)	(0.031)	(0.033)
Male Name	−0.009	−0.009	−0.021	−0.038
	(0.025)	(0.025)	(0.026)	(0.027)
No Name	0.086**	0.086**	0.083**	0.086**
	(0.030)	(0.030)	(0.031)	(0.033)
Constant	0.031†	0.031†	0.055**	0.094***
	(0.018)	(0.018)	(0.020)	(0.022)
N	135,587	135,587	135,587	135,587
Adjusted R^2	0.0006	0.0006	0.0005	0.0003

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included for the full sample but not reported. Robust errors are reported in the parentheses.

Automated Approaches

Finally, we also tried validating our results with another measure, toxicity, via PerspectiveAPI. PerspectiveAPI provides toxicity scores for comments on social media and news articles, among other things, using an algorithm to automatically classify texts based on their similarity to known toxic language. Our hope in using it was that we would be able to compare "apples to apples" across the two studies—a beneficial addition because they currently rely on different human coders. However, we found that the scores did not correlate well with any of our measures, perhaps because our snippets of text (SMS messages) are so short and thus subject to more measurement error.

For Study 1, PerspectiveAPI toxicity scores correlated with our coder-rated offensiveness measure at 0.56, silencing measure at 0.45, and withdrawal measure at 0.34. For Study 2, toxicity scores correlated with our coder-rated offensiveness measure at 0.62, discouragement measure at 0.61, silencing measure at 0.42, and withdrawal measure at 0.42. We were surprised to see that the withdrawal messages correlated so highly with the toxicity measure; we believe this is because the API is sensitive to edge cases (where the text may be mildly offensive but the volunteer(s) in question did not mark them as offensive): for instance, saying "that's dumb" or "you're an idiot" might not be marked as offensive by the volunteers, but they score highly with the API.

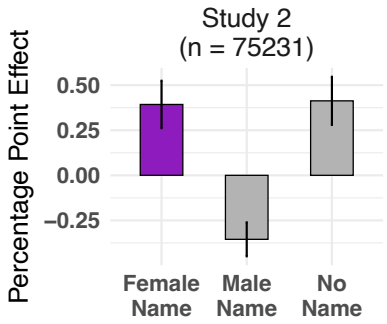
Ultimately, we were not confident that the toxicity scores would be informative because of their low correlations with our existing measures. Indeed, we relied on codings by human volunteers from the beginning due to pre-existing concerns about the difficulty of using automated classifiers (which typically measure similarity between two strings, and thus tend to be more reliable the longer the strings are). Accordingly, we found ourselves disappointed but unsurprised that we could not use PerspectiveAPI toxicity scores as a way to externally validate our volunteers' ratings.

Results for Discouragement Study 2 shows that respondents send more discouraging messages to volunteers using female names. Both female-named (two-tailed $p < .001$) and unnamed volunteers (two-tailed $p < .001$) were significantly more likely to receive discouraging replies than an ambiguously-named volunteer, while male-named volunteers were significantly less likely to receive discouraging replies (two-tailed $p < .001$). Overall, female-named volunteers can expect to receive 3.93 more discouraging messages per 1,000 messages than the ambiguously-named, and 7.48 more discouraging messages than male-named volunteers. Again, we

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

caution against over-emphasizing this result because of the reliability problems with this measure.

FIGURE S2. Mean Discouragement. Figure shows the average treatment effect by treatment condition with 95% confidence intervals estimated using OLS. The comparison category is the ambiguous name condition. Data is from Study 2 only.



Study Context, Delivery, and Timing

NGA is a national political action committee that focuses on engaging young people in the political process to support progressive causes. They typically organize on and around college campuses focusing on voter registration and voter turnout. They were originally founded in 2013 as NextGen Climate and primarily focused on environmental issues.

In 2018, we worked with NGA to conduct two experiments evaluating whether female volunteers receive more harassing SMS responses than male volunteers. NGA was interested in this experiment because they had received comments from their female volunteers and staff that they felt harassed when sending get-out-the-vote (GOTV) SMS messages to voters. The organization wanted to find new ways to maintain the same programmatic efficacy while ensuring that their volunteers and staff members felt safe while participating in the political process.

The first experiment was conducted during NGA's response to the Marjory Stoneman Douglas High School shooting in Florida. Activists were organizing an event, March for Our Lives, to protest the lack of gun control legislation. NGA texted their list of progressive voters to encourage them to participate in the protest. The organization has recruited for other protests in the past such as, the Women's March and the Climate March.

In the first experiment, the organization failed to deliver 18,457 SMS messages out of those individuals assigned to be part of the study because they lacked sufficient volunteers. The organization failed to deliver 36,290 SMS messages in the second experiment for the same reason.

In the second experiment, NGA texted progressive voters to encourage them to lobby their members of Congress to call for then-Environmental Protection Agency Director Scott Pruitt's removal. Pruitt had recently been accused of several ethical violations, which sparked intense lobbying for his removal by progressive organizations. NGA has recruited voters to lobby members of Congress in the past on several political issues.

Three days before the March for Our Lives protest in March, volunteers began sending messages to the first experiment's sample through Relay, an online SMS messaging platform. Later in April, volunteers began texting the second experiment's sample through Relay. Volunteers were not told about the study's intent to reduce potential demand effects. Additionally, only a select few staffers in the organization knew about the study in order to reduce the number of people who could induce demand effects.

Supporter Gender NGA does not keep data on the gender identity of individuals they talk to in their database, so we used the 'gender' package in R to code likely gender based on their first name. We were unable to predict the gender of a small number of supporters in both studies, and dropped these people from analyses examining

APSR Submission Template

the gender of the respondents. Dropping these individuals does not bias our estimates because we randomly assign treatment conditions for each individual, which means that those with "unknown" names are equally likely to be assigned each treatment condition.

TABLE S4. Supporter Gender by Study

Gender	Study 1	Study 2
Man	41.56%	38.97%
Unknown	5.99%	6.34%
Woman	52.45%	54.69%

TABLE S5. Supporter Gender by Condition, Study 1

Experimental Condition	Men	Unknown Gender	Women
Female-named	40.69%	5.83%	53.48%
Male-named	42.44%	6.23%	51.33%
Ambiguously-named	41.39%	5.95%	52.66%
Unnamed	41.7%	5.96%	52.35%

TABLE S6. Supporter Gender by Condition, Study 2

Experimental Condition	Men	Unknown Gender	Women
Female-named	38.33%	6.59%	55.08%
Male-named	39.63%	6.26%	54.1%
Ambiguously-named	38.7%	6.38%	54.91%
Unnamed	39.23%	6.1%	54.66%

Table S4 presents the sample for each study broken down by gender. Tables S5 and S6 do the same but break them down further by experimental condition. Both studies have more female than male supporters.

RESULTS AND ROBUSTNESS CHECKS

OLS Regressions for Results Reported Visually

The following tables provide the raw regression results for the analyses reported in Figures 1-3 in the paper. We first provide the pooled and individual study results for offensiveness, silencing, and withdrawal.

Offensiveness Table S7 is the basis for Figure 1. The first column pools both studies together using a binary measure. Respondents are .177 percentage points more likely to reply with an offensive message when a volunteer uses a female name compared to when they use an ambiguous name (two-tailed $p < .001$). Respondents are .097 percentage points less likely to respond with an offensive message when the volunteer uses a male name compared to when they use an ambiguous name (two-tailed $p < .01$). Respondents were .148 percentage points more likely to respond with an offensive message when the volunteer is unnamed (two-tailed $p < .001$). For context, volunteers using female names will receive 2.74 more offensive measures than volunteers using male

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

names. This basic pattern holds across both studies when analyzed individually using a dichotomized measure. Additionally, our main finding that respondents are more likely to respond with offensive messages when a volunteer uses a female name holds even when we use the original ordinal measure in Study 2.

TABLE S7. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving an Offensive Message (Dichotomized)

	Offensive Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.177***	0.151*	0.199***	0.202***	0.173*
	(0.042)	(0.060)	(0.059)	(0.051)	(0.075)
Male Name	−0.097**	−0.014	−0.164***	−0.053	−0.144*
	(0.031)	(0.049)	(0.039)	(0.036)	(.057)
No Name	0.148***	0.074	0.208***	0.139**	0.181*
	(0.041)	(0.055)	(0.059)	(.048)	(.075)
Constant	0.232***	0.191***	0.228***	0.156***	0.317***
	(0.028)	(0.035)	(0.035)	(0.032)	(0.053)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.0005	0.0001	0.0008	0.0005	0.0005

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included for the full sample but not reported. Robust standard errors are reported in the parentheses.

Silencing In both studies, we find that voters are more likely to silence female-named volunteers than all other named conditions. In Table S8, we report the OLS regression results with an added dummy variable indicating whether the data comes from the first or second experiment. In the first column, respondents were 0.117 percentage points more likely to respond to female-named volunteers than ambiguously-named volunteers (two-tailed $p < .001$). In contrast, volunteers assigned male names were 0.053 percentage points less likely to be silenced as the ambiguously-named volunteer (two-tailed $p = .034$). Unnamed volunteers were less likely to be silenced. The difference means that female-named volunteers are forced to end all future outreach for 8.36 more respondents out of every 1000 text messages they send to voters than male-named volunteers do.

Withdrawal Table S9 presents the OLS analysis whether the gender of the name affects the likelihood of a respondent politely ending all future activism from the organization. Respondents were 0.383 percentage points more likely to politely end all future outreach when volunteers used female names compared to the ambiguous name ($p < .001$). Respondents were 0.334 percentage points less likely to withdraw from outreach with volunteers using male names ($p < .001$). Across all samples, respondents were more likely to withdraw when contacted by female-named volunteers. Respondents were slightly less likely to withdraw with male-named volunteers. They were more likely to withdraw from all future outreach with unnamed volunteers but not significantly so.

Outcomes within Responses

Response Rates Table S10 shows the percentage of supporters in the sampling frame who responded to a message by experimental condition for each study. In both studies, respondents were more likely to respond to a

TABLE S8. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Silencing Message, Both Studies

	Silencing Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.117***	0.096*	0.134**	0.137***	0.110†
	(0.034)	(0.045)	(0.048)	(0.041)	(0.061)
Male Name	−0.053*	0.020	−0.111***	−0.026	−0.084†
	(0.025)	(0.039)	(0.032)	(0.029)	(0.047)
No Name	0.057†	0.040	0.069	0.039	0.079
	(0.031)	(0.041)	(0.045)	(0.034)	(0.058)
Constant	0.147***	0.106***	0.154***	0.094	0.217***
	(0.023)	(0.026)	(0.029)	(0.025)	(0.044)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.0002	0.00003	0.0004	0.0003	0.0002

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust standard errors are reported in the parentheses.

TABLE S9. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Withdrawal Message, Both Studies

	Withdrawal Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.383***	0.579***	0.227**	0.482***	0.323*
	(0.085)	(0.163)	(0.081)	(0.114)	(0.142)
Male Name	−0.334***	−0.313***	−0.350	−0.285**	−0.384**
	(0.072)	(0.144)	(0.059)	(0.095)	(0.121)
No Name	0.111	0.023	0.182*	0.164	0.010
	(0.081)	(0.151)	(0.080)	(0.107)	(0.133)
Constant	0.484***	1.754***	0.509***	0.499***	0.456***
	(0.055)	(0.107)	(0.052)	(0.073)	(0.090)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.004	0.0005	0.0010	0.0035	0.0061

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust standard errors are reported in the parentheses.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

volunteer using a female name than an ambiguous name. Respondents are less likely to respond to unnamed volunteers compared to volunteers using the ambiguous name. Respondents were less likely to respond to volunteers using male names in Study 1, but more likely to respond in Study 2 compared to volunteers using the ambiguous name. Table S13 presents the OLS regression results for the likelihood that a respondent responds given their volunteer-name assignment.

TABLE S10. Response Rate by Experimental Condition, Study 1 and Study 2

Experimental Condition	Study 1	Study 2
Female Name	7.55%	1.96%
Ambiguous Name	6.30%	1.17%
Male Name	5.47%	0.36%
No Name	6.63%	2.01%

TABLE S11. Response Rate by Experimental Condition and Respondent Gender, Study 1

Experimental Condition	Respondent Gender	Response Rates
Female Name	Men	7.99%
Female Name	Women	7.57%
Ambiguous Name	Men	7.2%
Ambiguous Name	Women	5.86%
Male Name	Men	5.81%
Male Name	Women	5.45%
No Name	Men	6.81%
No Name	Women	6.68%

TABLE S12. Response Rate by Experimental Condition and Respondent Gender, Study 2

Experimental Condition	Respondent Gender	Response Rates
Female Name	Men	2.06%
Female Name	Women	1.90%
Ambiguous Name	Men	1.22%
Ambiguous Name	Women	1.08%
Male Name	Men	0.36%
Male Name	Women	0.35%
No Name	Men	2.24%
No Name	Women	1.88%

Table S13 shows that female-named volunteers were more likely to receive responses than male-named and ambiguously-named volunteers. In the full sample, respondents were .991 percentage points more likely to respond to a female-named volunteer (two-tailed $p < .001$). Respondents were 0.818 percentage points less likely to respond to a male-named volunteer as an ambiguously-named volunteer ($p < .001$), and they were .610 percentage points more likely to respond to an unnamed volunteer as an ambiguously-named volunteer ($p < .001$). Contextually, this means that a female-named volunteer would receive 18.09 more replies than a male-named volunteer if they both were to message 1,000 respondents. We see the same pattern across all samples.

Count of Responses and Response Types by Volunteer

In Tables S14-S15, we show the average raw count of the number of responses, offensive responses, silencing responses, and withdrawal responses per volunteer, per treatment condition, in each study, along with percentages for the types of responses divided by the average number of responses.

APSR Submission Template

TABLE S13. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Response, Both Studies

	Responses				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.991***	1.250***	0.785***	1.205***	0.815***
	(0.148)	(0.293)	(0.128)	(0.197)	(0.247)
Male Name	−0.818***	−0.825**	−0.812***	−0.592***	−1.104
	(0.130)	(0.270)	(0.090)	(0.173)	(0.217)
No Name	0.610***	0.328	0.837***	0.804***	0.366
	(0.145)	(0.283)	(0.129)	(0.192)	(0.240)
Constant	1.180***	6.299***	1.172***	0.948***	1.453***
	(0.094)	(0.197)	(0.078)	(0.124)	(0.157)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.0196	0.0009	0.0033	0.0201	0.0207

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust standard errors are reported in the parentheses.

As with the regression results reported in Tables S16-S18, we caution the reader from over-interpreting these tables because they are vulnerable to post-treatment bias. We are implicitly conditioning on a post-treatment variable, responding, by comparing the number of responses to the number for each response type.

TABLE S14. Count of Responses and Response Types by Volunteer for Study 1

Condition	Responses	Offensive	Silencing	Withdrawal
Ambiguous Name	73.5	2.23 (3%)	1.23 (1.6%)	20.5 (27.9%)
Female Name	93.8	4.25 (4.5%)	2.5 (2.7%)	29 (30.9%)
Male Name	52	1.69 (3.3%)	1.19 (2.3%)	13.7 (26.3%)
No Name	38.4	1.54 (4%)	0.85 (2.2%)	10.3 (26.8%)

TABLE S15. Count of Responses and Response Types by Volunteer for Study 2

Condition	Responses	Offensive	Silencing	Withdrawal
Ambiguous Name	22.1	1.18 (5.3%)	2.9 (13.1%)	9.6 (43.4%)
Female Name	40.8	2.14 (5.2%)	6 (14.7%)	15.3 (37.5%)
Male Name	17	1.91 (11.2%)	2 (11.8%)	7.5 (44.1%)
No Name	37.8	2.95 (7.8%)	4.2 (11.1%)	13 (34.4%)

Regression Analyses Conditional on Receiving a Response The following tables describe how respondents responded to differently named volunteers conditional on responding in the first place. We refrain from providing these results in the main text because they are vulnerable to post-treatment bias: we know that people are already more likely to respond to female-named volunteers than male-named volunteers. Therefore, the only way we can interpret these results is if the sequential ignorability assumption holds, which is both unlikely and unverifiable. However, we provide these analyses to provide suggestive evidence that our results are not purely a function of

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

the fact that female-named volunteers are more likely to receive responses. These analyses require us to drop a substantial portion of our sample, which will inflate our standard errors.

Offensiveness

Table S16 shows the dichotomized OLS offensive reply analysis conditional on receiving a reply. The first column displays the effect of a gendered name on the likelihood of receiving an offensive reply conditional on receiving a reply for the full sample. Female-named volunteers were 1.668 percentage points more likely to receive offensive replies. Male-named volunteers were 0.121 percentage points more likely to receive offensive replies. Unnamed volunteers were 1.266 percentage points more likely to receive offensive replies. Only the female-named volunteer condition was statistically significant (two-tailed ($p < .1$). The women-only and men-only sample results are consistent with the full sample results.

TABLE S16. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Offensive Message Conditional on Receiving a Response (Dichotomized)

	Offensive Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	1.668†	1.493†	2.341	2.655*	0.994
	(0.976)	(0.832)	(3.432)	(1.242)	(1.567)
Male Name	0.121	0.209	−1.810	0.621	−0.193
	(0.943)	(0.828)	(5.345)	(1.184)	(1.510)
No Name	1.266	0.967	2.236	1.536	1.258
	(0.997)	(0.833)	(3.410)	(1.242)	(1.635)
Constant	19.924***	3.037***	19.457***	15.942***	24.360***
	(1.408)	(0.556)	(2.668)	(1.780)	(2.354)
N	4,946	3,912	1,034	2,561	2,169
Adjusted R^2	0.072	0.0003	−0.002	0.061	0.084

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust errors are reported in the parentheses.

Silencing

Table S17 shows the OLS silencing reply analysis conditional on receiving a reply. The first column presents the effect of a gendered name on the likelihood of silencing conditional on receiving a reply for the full sample. Female-named volunteers are 1.185 percentage points more likely to receive silencing replies. Male-named volunteers were 0.367 percentage points more likely to receive a silencing reply. Unnamed volunteers were 0.099 percentage points less likely to receive silencing replies. For context, if a female-named volunteer were to receive 100 replies, they would receive 1.185 more silencing replies. These effects are in the same direction as those reported in the paper, but are rarely statistically significant. All columns show a similar pattern.

Withdrawal

Table S18 shows the OLS silencing reply analysis conditional on receiving a reply. The first column presents the effect of a gendered name on the likelihood of withdrawal conditional on receiving a reply for the full sample. Female-named volunteers are 0.947 percentage points more likely to receive withdrawal replies. Male-named volunteers were 2.819 percentage points less likely to receive a silencing reply. Unnamed volunteers were 2.857 percentage points less likely to receive silencing replies. For context, if a female-named volunteer were to receive 100 replies, they would receive .947 more silencing replies. As above, the effects follow a similar pattern to those reported in the main paper, but generally are not statistically significant. All columns show a similar pattern.

TABLE S17. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Silencing Message Conditional on Receiving a Response

	Silencing Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	1.185	0.989	1.592	1.941†	0.552
	(0.796)	(0.635)	(2.934)	(1.006)	(1.300)
Male Name	0.367	0.608	−1.357	0.570	0.276
	(0.776)	(0.664)	(4.582)	(0.972)	(1.256)
No Name	−0.099	0.527	−2.011	−0.179	−0.144
	(0.784)	(0.623)	(2.793)	(0.936)	(1.328)
Constant	12.454***	1.675***	13.122***	9.373***	16.398***
	(1.167)	(0.416)	(2.276)	(1.400)	(2.054)
N	4,946	3,912	1,034	2,561	2,169
Adjusted R^2	0.044	−0.0002	−0.0007	0.0336	0.0576

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust standard errors are reported in the parentheses.

TABLE S18. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Withdrawal Message Conditional on Receiving a Response

	Withdrawal Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.947	3.011	−7.019†	1.607	0.693
	(1.802)	(1.994)	(4.175)	(2.506)	(2.713)
Male Name	−2.819	−1.698	−6.222	−2.762	−2.017
	(1.992)	(2.097)	(6.742)	(2.771)	(3.003)
No Name	−2.857	−1.013	−9.918*	−3.727	−2.586
	(1.809)	(2.011)	(4.124)	(2.510)	(2.725)
Constant	37.354***	27.539***	42.986***	44.141***	29.787***
	(1.950)	(1.446)	(3.337)	(2.754)	(2.918)
N	4,946	3,912	1,034	2,561	2,169
Adjusted R^2	0.007	0.001	0.003	0.024	−0.001

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust standard errors are reported in the parentheses.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

Other Outcomes In the sections that follow, we provide the regression results for other outcomes referenced in the main manuscript.

March Attendance

Table S19 presents the OLS analysis whether the gender of the name affects the likelihood of a respondents committing to attending the march. In the full sample, respondents were more likely to tell female-named volunteers that they would attend the march ($p = .058$). Respondents were slightly less likely to tell male-named and unnamed volunteers that they would attend the march, The pattern holds in the next two columns, except with unnamed volunteers in the women-only sample. All of these coefficients are statistically insignificant.

TABLE S19. Effect of a Volunteer's Perceived Gender on the Likelihood of March Attendance, Using Self-Reports

	Reported March Attendance		
	Full sample	Women only	Men only
Female Name	0.247†	0.452*	0.017
	(0.131)	(0.198)	(0.177)
Male Name	−0.154	−0.122	−0.182
	(0.120)	(0.181)	(0.166)
No Name	−0.013	0.041	−0.081
	(0.124)	(0.186)	(0.171)
Constant	1.174***	1.365***	0.972**
	(0.087)	(0.130)	(0.124)
N	60,356	31,656	25,083
Adjusted R^2	0.0001	0.0002	-0.0001

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Data derived from Study 1. Baseline category is the ambiguous name condition. Robust standard errors are reported in the parentheses.

Constituent Calls

Table S20 displays the effect of a gendered name on a respondent's likelihood of reporting that they would call their congressperson. Across all three samples, respondents were more likely to commit to calling their congressperson when a female-named or unnamed volunteer contacted them. The effects were only statistically significant in the full sample and the men-only sample. Male-named volunteers were less likely to receive a call commitment, but these effects were statistically insignificant.

Gendered Insults

Below, we describe the proportion of times a respondent used a gendered insult such as: "slut," "whore," "cunt," or "bitch." Table S21 shows the proportions for Study 1 and Study 2. The proportions are low across all cases.

Outcomes Moderated by Volunteer Gender

In this section, we evaluate how the effects vary by the real-life gender of the volunteer. We caution readers from over-interpreting the following results because they are subject to post-treatment bias. Recall that NGA randomized what names volunteers used to contact each supporter, but that each volunteer would be assigned a batch of supporters. Randomization is therefore at the level of the supporter, not the volunteer. Additionally, there are very few men volunteers. There were 27 women volunteers and seven men volunteers in Study 1. There

TABLE S20. Effect of a Volunteer's Perceived Gender on the Likelihood of Calling Their Representative, Using Self-Reports

	Reported Calling		
	Full sample	Women only	Men only
Female Name	0.177***	0.107	0.267**
	(0.053)	(0.068)	(0.089)
Male Name	−0.106**	−0.125*	−0.070
	(0.036)	(0.048)	(0.056)
No Name	0.176***	0.147*	0.242**
	(0.053)	(0.071)	(0.086)
Constant	0.175***	0.184***	0.151***
	(0.030)	(0.042)	(0.045)
N	75,231	41,143	29,321
Adjusted R^2	0.0006	0.0004	0.0007

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Data derived from Study 2. Baseline category is the ambiguous name condition. Robust standard errors are reported in the parentheses.

TABLE S21. Gendered Results, Study 1 and Study 2

Gender	Study 1	Study 2
Female Name	0.02%	0%
Ambiguous Name	0.007%	0%
Male Name	0.013%	0.005%
No Name	0%	0.011%

APSR Submission Template

were 24 women volunteers and four men volunteers in Study 2. As a reminder, two women raters coded our original offensiveness (and discouragingness) ratings. We then asked an additional three women undergraduate research assistants to code offensiveness.

We investigate whether the results using volunteer self-reported offensiveness vary by volunteer gender in Study 1 because in Study 2 the two volunteers who coded all responses were both women. We run an F-test comparing our basic model against a model including an interaction for volunteer gender with every treatment variable. We regress offensiveness on our female name, male name, and no name dummy variables in our basic model. We regress offensiveness on our female name, male name, no name dummy variables, a texter gender dummy variable where woman equals 1, and all of the interactions between the experimental condition dummies and the texter gender dummy variable. The F-statistic is 1.001 ($p = .864$), which is not statistically significant. This suggests women and men volunteers did not code offensive messages differently in Study 1.

Logistic Regressions

Below, we replicate our original ordinary least squares analyses using logistic regression. Our results replicate the plots in the main text and the regression tables in the Supplementary Materials above.

Offensiveness Table S22 replicates Table S7 using logistic regression. The first column shows that female-named and unnamed volunteers were significantly likelier than ambiguously-named volunteers to receive offensive replies. Male-named volunteers were significantly less likely to receive offensive replies than ambiguously-named volunteers. The second and third columns repeat the analysis from before for the first and second study separately. The fourth and fifth columns implement the same analysis for the women-only and men-only samples, respectively. They all show the same pattern where female-named and unnamed volunteers are more likely to receive offensive replies, while male-named volunteers are less likely to receive offensive replies.

TABLE S22. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving an Offensive Message (Dichotomized)

	Offensive Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.610^{***}	0.582^{*}	0.628^{***}	0.889^{***}	0.463^{*}
	(0.147)	(0.233)	(0.189)	(0.233)	(0.203)
Male Name	-0.615^{**}	-0.074	-1.278^{***}	-0.468	-0.671^{*}
	(0.199)	(0.268)	(0.327)	(0.318)	(.27)
No Name	0.532^{***}	0.328	0.649^{***}	0.684^{**}	0.48^{*}
	(0.149)	(0.244)	(0.189)	(.241)	(.201)
Constant	-6.084^{***}	6.257^{***}	-6.08^{***}	-6.492^{***}	-5.761^{***}
	(0.126)	(0.186)	(0.153)	(0.208)	(0.171)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.014	0.005	0.025	0.018	0.013

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included for the full sample but not reported. Robust errors are reported in the parentheses.

Silencing Table S23 replicates Table S8 using logistic regression. The first column shows that female-named volunteers were significantly likelier than ambiguously-named volunteers to receive silencing replies. Male-

named volunteers were significantly less likely to receive silencing replies than ambiguously-named volunteers. Unnamed volunteers were more likely to receive silencing replies than ambiguously-named volunteers, but the effect was not statistically significant at conventional levels. The second and third columns repeat the analysis from before for the first and second study separately. The fourth and fifth columns implement the same analysis for the women-only and men-only samples, respectively. They all show the same basic pattern where female-named volunteers are more likely to receive silencing replies, while male-named volunteers are less likely to receive silencing replies. Unnamed volunteers were more likely to receive silencing replies, but the statistical significance of the effect varied across samples.

TABLE S23. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Silencing Message, Both Studies with Logistic Regression

	Silencing Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.635***	0.646**	0.628	0.944***	0.454†
	(0.185)	(0.310)	(0.230)	(0.295)	(0.252)
Male Name	−0.512*	0.169	−1.289***	−0.357	−0.576†
	(0.244)	(0.340)	(0.399)	(0.392)	(0.324)
No Name	0.356†	0.325	0.373	0.372	0.346
	(0.195)	(0.329)	(0.242)	(0.326)	(0.256)
Constant	−6.541***	−6.853***	−6.475***	−6.993***	−6.146***
	(0.161)	(0.250)	(0.186)	(0.263)	(0.216)
N	135,587	60,356	75,231	72,799	54,404
R^2	0.011	0.004	0.021	0.015	0.010

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust standard errors are reported in the parentheses.

Withdrawal Table S24 replicates Table S9 using logistic regression. The first column shows that female-named volunteers were likelier than ambiguously-named volunteers to receive withdrawal replies. Male-named volunteers were significantly less likely to receive withdrawal replies than ambiguously-named volunteers. Unnamed volunteers were more likely to receive withdrawal replies than ambiguously-named volunteers, but the effect was not statistically significant. The second and third columns repeat the analysis from before for the first and second study separately. The fourth and fifth columns implement the same analysis for the women-only and men-only samples, respectively. They all show the same pattern where female-named volunteers are more likely to receive withdrawal replies, while male-named volunteers are less likely to receive withdrawal replies. Unnamed volunteers were more likely to receive withdrawal replies, but the statistical significance of the effect varied across samples.

OLS Regressions Using Cluster-Robust Standard Errors

The following tables replicate our key results using cluster-robust standard errors, clustering at the volunteer level. We refrain from providing these regressions in the main text because the randomization occurred at the respondent level, so robust standard errors should be sufficient. However, there might be concerns that volunteers change as they interact with more respondents across experimental conditions. Note that the standard errors might have an anti-conservative bias in the study-specific regressions because of how few volunteers there are in each study. We first provide the pooled and individual study results for offensiveness, silencing, and withdrawal.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

TABLE S24. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Withdrawal Message, Both Studies with Logistic Regression

	Withdrawal Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.313***	0.291***	0.371**	0.407***	0.242*
	(0.070)	(0.082)	(0.133)	(0.097)	(0.106)
Male Name	−0.381***	−0.200*	−1.166***	−0.350**	−0.385**
	(0.083)	(0.092)	(0.209)	(0.117)	(0.122)
No Name	0.101	0.013	0.307*	0.157	0.009
	(0.073)	(0.087)	(0.135)	(0.103)	(0.111)
Constant	−5.285***	−4.025***	−5.274***	−5.217***	−5.409***
	(0.069)	(0.062)	(0.102)	(0.092)	(0.112)
N	135,587	60,356	75,231	72,799	54,404
R^2	0.04	0.003	0.02	0.03	0.05

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust standard errors are reported in the parentheses.

Offensiveness Table S25 replicates Table S7 with cluster-robust standard errors. Our results match our original findings. Respondents were more likely to send offensive replies to volunteers using female names or no names. Additionally, respondents were less likely to send offensive replies to volunteers using male names. The general pattern of results occurs in all columns.

TABLE S25. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving an Offensive Message (Dichotomized and Cluster-Robust SEs)

	Offensive Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.177**	0.151*	0.199*	0.202**	0.173*
	(0.057)	(0.070)	(0.084)	(0.062)	(0.075)
Male Name	−0.097†	−0.014	−0.164	−0.053	−0.144†
	(0.052)	(0.067)	(0.109)	(0.044)	(.080)
No Name	0.148*	0.074	0.208*	0.139*	0.181†
	(0.073)	(0.100)	(0.090)	(.060)	(.105)
Constant	0.232†	0.191*	0.228	0.156	0.317†
	(0.135)	(0.082)	(0.149)	(0.097)	(0.182)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.0005	0.0001	0.0008	0.0005	0.0005

$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included for the full sample but not reported. Cluster-robust standard errors are reported in the parentheses.

APSR Submission Template

Silencing Table S26 replicates Table S8 with cluster-robust standard errors. Our results match our original findings. Respondents were more likely to send silencing replies to volunteers using female names or no names. Additionally, respondents were less likely to send silencing replies to volunteers using male names. The general pattern of results occurs in all columns.

TABLE S26. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Silencing Message, Both Studies (Cluster-Robust SEs)

	Silencing Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.117**	0.096	0.134*	0.137***	0.110†
	(0.040)	(0.059)	(0.053)	(0.041)	(0.061)
Male Name	−0.053	0.020	−0.111	−0.026	−0.084
	(0.036)	(0.040)	(0.076)	(0.031)	(0.060)
No Name	0.057	0.040	0.069	0.039	0.079
	(0.053)	(0.064)	(0.078)	(0.039)	(0.077)
Constant	0.147†	0.106*	0.154	0.094	0.217†
	(0.087)	(0.041)	(0.101)	(0.058)	(0.129)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.0002	0.00003	0.0004	0.0002	0.0002

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Cluster-robust standard errors are reported in the parentheses.

Withdrawal Table S27 replicates Table S9 with cluster-robust standard errors. Our results match our original findings, but are noisier. Respondents were more likely to politely end activist outreach with volunteers using female names. Additionally, respondents were more likely to withdraw with unnamed volunteers and less likely to withdraw with volunteers using male names, but these were never statistically significant. The general pattern of results occurs in all columns.

Instrumental Variable Regressions

In the following section, we provide estimates of the average treatment effect on the treated, using instrumental variable regressions to correct for one-sided noncompliance. Recall that NGA could not ultimately send several thousand planned messages due to volunteers running out of time; these estimates exclude unsent messages, so the effect sizes are generally much larger than those we report in the main paper.

Offensiveness Table S28 shows the IV analysis for how respondents replies to differently gendered names using the dichotomized offensive ratings from both studies. The first column shows that female-named and unnamed volunteers were significantly likelier than ambiguously-named volunteers to receive offensive replies. Male-named volunteers were significantly less likely to receive offensive replies than ambiguously-named volunteers. The second and third columns repeat the analysis from before with women-only and men-only samples, respectively. They both show the same pattern where female-named and unnamed volunteers are significantly more likely to receive offensive replies, while male-named volunteers are significantly less likely to receive offensive replies.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

TABLE S27. Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Withdrawal Message, Both Studies (Cluster-Robust SEs)

	Withdrawal Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.383*	0.579	0.227	0.482**	0.323
	(0.174)	(0.369)	(0.166)	(0.170)	(0.222)
Male Name	−0.334†	−0.313	−0.350	−0.285†	−0.384†
	(0.176)	(0.300)	(0.217)	(0.164)	(0.221)
No Name	0.111	0.023	0.182	0.164	0.010
	(0.298)	(0.618)	(0.187)	(0.275)	(0.359)
Constant	0.484†	1.754**	0.509	0.499†	0.456†
	(0.270)	(0.584)	(0.310)	(0.281)	(0.265)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.004	0.0005	0.0010	0.0035	0.0061

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Cluster-robust standard errors are reported in the parentheses.

TABLE S28. Complier Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving an Offensive Message (Dichtomized)

	Offensive Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.240***	0.201*	0.271***	0.272***	0.235*
	(0.057)	(0.079)	(0.081)	(0.069)	(0.102)
Male Name	−0.287**	−0.023	−1.237***	−0.160	−0.412*
	(0.091)	(0.083)	(0.297)	(0.108)	(0.164)
No Name	0.191***	0.097	0.266***	0.179**	0.233*
	(0.053)	(0.072)	(0.076)	(0.062)	(0.097)
Constant	0.217***	0.191	0.228***	0.149***	0.294***
	(0.025)	(0.035)	(0.035)	(0.028)	(0.046)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.0004	0.0005	−0.0025	0.0008	0.0002

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust errors are reported in the parentheses.

APSR Submission Template

Silencing Table S29 replicates the analysis done in Table S8 but corrects for non-compliance with IV regression. We find a similar pattern here.

TABLE S29. Complier Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Silencing Message, Both Studies by Study

	Silencing Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.158***	0.127*	0.183**	0.184***	0.149†
	(0.045)	(0.060)	(0.066)	(0.055)	(0.082)
Male Name	−0.156*	0.033	0.838***	−0.079	−0.240†
	(0.074)	(0.066)	(0.244)	(0.086)	(0.134)
No Name	0.073†	0.053	0.089	0.051	0.102
	(0.040)	(0.053)	(.057)	(0.044)	(0.075)
Constant	0.139***	0.106***	0.154***	0.090***	0.203***
	(0.020)	(0.026)	(0.029)	(0.022)	(0.039)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.0002	0.0003	0.0004	0.0004	−.00001

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included for the model using both studies. Robust standard errors are reported in the parentheses.

Withdrawal Table S30 shows the IV withdrawal analysis for both studies. Our findings replicate the same pattern found in Table S9.

TABLE S30. Complier Effect of a Volunteer's Perceived Gender on the Likelihood of Receiving a Withdrawal Message, Both Studies

	Withdrawal Replies				
	Full Sample	Study 1	Study 2	Women only	Men only
Female Name	0.517***	0.770***	0.310**	0.649***	0.440*
	(0.115)	(0.217)	(0.111)	(0.154)	(0.192)
Male Name	−0.986***	−0.528*	−2.634***	−0.863**	−1.102**
	(0.214)	(0.243)	(0.452)	(0.289)	(0.349)
No Name	0.144	0.030	0.232*	0.212	0.014
	(0.104)	(0.198)	(0.102)	(0.138)	(0.171)
Constant	0.434***	1.754***	0.509***	0.459***	0.395***
	(0.047)	(0.107)	(0.052)	(0.063)	(0.076)
N	135,587	60,356	75,231	72,799	54,404
Adjusted R^2	0.003	0.001	0.001	0.003	0.005

†p < .1; *p < .05; **p < .01; ***p < .001

Baseline category is the ambiguous name condition. Experiment-level fixed effects are included. Robust standard errors are reported in the parentheses.

APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template