Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Private Preferences, Public Expression, and Campaign Persuasion: Experimental Evidence from the 2024 U.S. Presidential Election

Laura Paler

lpaler@american.edu

American University

Sarah Khan

American University

Research Article

Keywords: campaign persuasion, political expression, preference falsification, social norms, political preferences, gender politics

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11391749/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Private Preferences, Public Expression, and Campaign Persuasion: Experimental Evidence from the 2024 U.S. Presidential Election

October 15, 2026

Abstract

Research on electoral and campaign persuasion typically examines how appeals change what voters think, but appeals can also change what voters are willing to express to others. Because public expression depends on social incentives as well as personal beliefs, campaign appeals can affect private preferences and public expression differently, creating, widening, or narrowing gaps between them. Existing research rarely examines these outcomes together however. We address this by studying a campaign ad targeting women during the 2024 U.S. presidential election using experiments embedded in a pre-election survey of 2,056 primarily Republican and Independent women. By cross-randomizing ad exposure and whether respondents express their candidate preferences privately in an online survey or publicly to a co-partisan virtual enumerator, we identify the ad's effects in each domain. The ad reduced private Trump support significantly more than support expressed to a female co-partisan. We examine whether this asymmetry reflects social costs among Republican women, as anticipated by the "shy voter" account. Instead, the pattern is most evident among Independent women and is consistent with social benefits: these women initially expressed less Trump support publicly than privately, and the ad brought their private preferences into closer alignment with their public expression. These findings show how the same private effect without a corresponding public effect can have different interpretations depending on the social environment. Studying both private preferences and public expression provides a fuller account of what persuasion changes and of its potential—and limits—as a source of electoral change.

Word Count: 8,833

Keywords: campaign persuasion, political expression, preference falsification, social norms, political preferences, gender politics

Statements and Declarations

Funding. This research was supported by university research funds and received no external funding.

Competing Interests. The author declares no financial or non-financial competing interests.

Pre-Registration. This study was pre-registered on OSF. An anonymized version of the pre-analysis plan accompanies the manuscript and is referenced in the manuscript in footnote 9. Deviations from the pre-registration are discussed in Appendix K.

Data Availability. All data and script files needed to replicate the published results will be made publicly available on Dataverse upon acceptance of the manuscript.

Compliance with Ethical Standards. This study was approved by the authors' university Institutional Review Board and was conducted in accordance with the APSA guidelines on ethical standards. All participants provided informed consent to participate and were informed that their confidential responses would be combined with those of other participants and published as aggregate results, and that the research team could use and publish data

containing no information that would reveal their identities. The consent script and a discussion of research ethics appear in Appendix J.

Supplementary Materials. Supplementary materials containing the survey questions, stimuli, and procedures accompany the manuscript as a separate file and are referenced throughout the text.

Use of Generative AI. Generative AI (ChatGPT 6.1 and Claude Opus 5.5 as well as earlier models) was used to assist with drafting and revising text, organizing and clarifying theoretical arguments, and presenting and interpreting results, including producing figures from regression code supplied by the authors. The authors reviewed and verified all AI-assisted outputs and take responsibility for the manuscript's content. This statement is also included in the main text after the Conclusion section.

1 Introduction

A vast literature on electoral and campaign persuasion examines how appeals change voters' preferences, attitudes, candidate evaluations, and vote intentions (Huber and Arceneaux, 2007; Gerber et al., 2011; Hewitt et al., 2024; Coppock, Hill and Vavreck, 2020; Kalla and Broockman, 2018; Broockman and Kalla, 2023). These outcomes are understood to reflect voters' underlying political views, which can shape vote choice and other political behaviors (Druckman, 2022). From this perspective, a campaign message succeeds to the extent that it shifts those underlying views in favor of a candidate.

Yet campaign persuasion can also influence what individuals are willing to express publicly. Public expression—the outward display of attitudes or preferences—can amplify the effects of a campaign message through political discussion, interpersonal persuasion, social media sharing, and other visible forms of political action (Druckman, Levendusky and McLain, 2018; Carlson, 2019). Campaign ads may encourage such expression by conveying social information: an ad that depicts people expressing support for or opposition to a candidate signals both what others are doing and what they regard as acceptable or desirable (Cialdini, Reno and Kallgren, 1990). By shifting perceived norms, these signals can reduce the social costs or increase the social benefits of expressing a view (Bursztyn, Egorov and Fiorin, 2020; Carlson and Settle, 2016; Suhay, 2015). As more people express that view, they in turn signal its acceptability to others, creating the potential for broader political momentum (Noelle-Neumann, 1974; Mutz, 1992; Bond et al., 2012; Valentim, 2024).

Despite the importance of private preferences and public expression as distinct outcomes of campaign persuasion, existing research rarely examines them together. This matters because these outcomes need not move together. Changing private preferences involves changing personal beliefs—how someone thinks or feels about a candidate—while public expression additionally depends on the social incentives associated with communicating those beliefs to others (Kuran, 1995; Blair, Coppock and Moor, 2020; Levitan and Verhulst, 2016). Campaign ads may change personal beliefs and social incentives to different degrees, producing effects on both private preferences and public expression, neither, or larger effects on one than the other. Investigating these patterns is critical to understanding the effects of persuasive appeals.

These public–private effect patterns also matter for understanding campaign persuasion as a

source of electoral change. Consider an ad that moves an individual's private preferences toward a candidate but has no effect on public expression. This pattern can have different interpretations depending on the social environment. When individuals face social costs—disapproval or sanctions for publicly supporting the candidate—it could reflect a private change that does not carry over into public expression, indicating limited change. In contrast, when public support is socially rewarded, individuals may already express support they do not privately hold (Blinder, Ford and Ivarsflaten, 2013; Willer, Kuwabara and Macy, 2009; Pickup, Kimbrough and de Rooij, 2022). The ad could then narrow this pre-existing gap, bringing private preferences, and potentially private behavior, into closer alignment with what individuals already express publicly. Thus, interpreting the private and public effects of campaign persuasion requires understanding the social incentives individuals face and how an appeal reshapes the relationship between private preferences and public expression.

We examine these questions in the context of campaign persuasion targeting Republican and Independent women during the 2024 U.S. presidential election, a setting in which commentators and campaign strategists openly speculated about divergences between women's private political preferences and public expression (Washington Post, 2024; CNN, 2024). Specifically, we study a campaign ad produced and funded by Vote Common Good and released in the final days of the campaign, encouraging these women to defect from Trump and support Harris. It depicted a woman who arrives at a polling place with her Trump-supporting husband, votes for Harris while concealing her choice from him, and exchanges a knowing glance with another woman.[1] By emphasizing the stakes of the election and encouraging women to make their own choice, the ad may have shifted private preferences away from Trump and toward Harris. By portraying women signaling shared opposition to Trump to one another, it may also have changed the perceived social incentives for expressing that opposition. The ad thus provides an opportunity for examining how a campaign appeal affects private preferences and public expression, especially to other co-partisan women.

The electoral context also highlights the importance of examining whether these effects widen or narrow the gap between private preferences and public expression. The ad aired amid widespread speculation about "shy" women voters: Republican women who privately opposed Trump but were reluctant to say so publicly due to perceived social costs (NPR, 2024). The ad could have

[1]The ad can be viewed here: https://www.youtube.com/watch?v=FaCPck2qDhk.

increased the pool of such voters by shifting private preferences while social costs continued to deter public expression, even if the ad reduced those costs. This reasoning motivated our preregistered hypotheses: that the ad would reduce private support for Trump and increase private support for Harris, with larger effects on private preferences than on public expression.

We test these hypotheses in an online survey of 2,056 primarily Republican and Independent women conducted immediately before the election. Respondents were randomly assigned to watch the ad or to a pure control condition. To distinguish effects on private preferences and public expression, we cross-randomized respondents to a *private* condition, in which they expressed their views in a standard online survey, or a *public* condition, in which they expressed their views to a virtual female or male copartisan enumerator. This design builds on evidence that anonymous online surveys better capture private preferences and that socially relevant audiences shape expressed responses (Tourangeau and Yan, 2007; Huddy et al., 1997; Krysan and Couper, 2003; Adida et al., 2016; Blair, Coppock and Moor, 2020; Bursztyn et al., 2025). It captures a key feature of public expression: communicating attitudes to a potentially socially relevant other. Including male enumerators allows us to examine whether the ad's effects on public expression depend on the audience's gender, as we would expect if its portrayal of shared opposition among women changes the social incentives for expressing opposition to Trump to other women. Respondents expressed their views in their assigned condition both before and after ad exposure, allowing us to examine whether they initially expressed different levels of candidate support privately and publicly, whether the ad affected each domain differently, and whether those effects widened or narrowed the gap between private preferences and public expression.

Consistent with the direction of our preregistered predictions, the ad reduced private support for Trump by 0.10 standard deviations, and this effect was significantly larger than its effect on expression to the female enumerator, with a public–private difference of 0.17 standard deviations. The effects on Harris support point in the predicted direction but are smaller and not statistically significant; the ad also shifted private support toward third-party candidates. Although the pattern emerges primarily for one candidate, the results show that a campaign ad can change private preferences more than public expression.

We then examine whether this pattern reflects the social costs anticipated by the "shy voter" account, which should be most pronounced among Republican women. We find no detectable ad

effects on their privately held or publicly expressed Trump support. Instead, effects were most clearly evident among Independent women, where the results point to social benefits rather than social costs. Without the ad, Independent women already expressed more opposition to Trump to a female copartisan than they held privately, and the ad reduced their private support without a detectable additional effect on public expression. The ad's primary effect was thus to bring private preferences into closer alignment with already elevated public opposition to Trump. Differences between private and public expression before ad exposure, together with the ad's effects on Independent women's perceptions of Harris support among other Independent women, provide further evidence consistent with this social benefits interpretation.

This paper contributes to research on both campaign persuasion and social influences on political expression. First, it shows why studying private preferences and public expression together matters for understanding how campaign appeals work. Research on campaign persuasion examines how appeals change voters' preferences and how their effects spread through social networks (Druckman, Levendusky and McLain, 2018; Druckman, 2022), while research on political expression examines how social incentives shape what people express (Carlson and Settle, 2016; Levitan and Verhulst, 2016). Bringing these literatures together raises a question rarely examined by either: whether an appeal can change what voters privately prefer, what they express to others, or both. The answer matters because the two outcomes play distinct roles in elections. Private preferences can inform vote choice and other political behaviors perceived as private, whereas public expression shapes what others learn about prevailing views and can extend a message's influence beyond those who encounter it directly.[2] Studying either outcome alone can therefore give an incomplete account of what campaign persuasion changes.

Second, the paper connects campaign persuasion to the micro-foundations of social change by examining how appeals reshape the relationship between private preferences and public expression. Large literatures examine how social pressure produces gaps between the two through concealment or public conformity—forms of preference falsification in which people hide or misrepresent their private preferences in response to perceived social pressures (Kuran, 1995; Hayes, Scheufele and

[2]Voting may not always be perceived as private (Gerber et al., 2013), in which case social incentives may also shape how private preferences translate into vote choice.

Huge, 2006; Carlson and Settle, 2016; Levitan and Verhulst, 2016; Willer, Kuwabara and Macy, 2009; Blinder, Ford and Ivarsflaten, 2013; Pickup, Kimbrough and de Rooij, 2022; Valentim, 2024). These literatures also identify how such gaps change. Bringing concealed views into public can reveal that support is more widespread than believed, prompting further expression and change (Kuran, 1995; Prentice and Miller, 1993; Bursztyn, González and Yanagizawa-Drott, 2020). Alternatively, social change can begin with shifts in public expression: changes in perceived norms may lead people to express views they do not yet hold privately, a form of norm compliance (Paluck, 2009). As people internalize these norms, private preferences may move toward public expression, narrowing the gap (Tankard and Paluck, 2016) and making public expression less dependent on social rewards and more durable (Kelman, 1958, 2006).

Less attention has been paid to how campaign persuasion itself creates, deepens, or reduces these gaps. Importantly, the same pattern of effects can widen or narrow these gaps. A larger effect on private preferences than on public expression could deepen preference falsification where social costs prevail, yet in our setting it narrowed an existing gap among Independent women, consistent with a shift from norm compliance toward internalization. Campaign persuasion can therefore affect not only what voters privately prefer and publicly express, but whether the two move closer together or further apart.

We do not expect this specific pattern to recur in every setting. Which pattern of effects a campaign appeal produces depends on whether it changes personal beliefs, social incentives, or both, as well as on the social environment an individual faces. Studying private and public outcomes together, alongside the social incentives that discourage or encourage public expression, can clarify both the potential and the limits of campaign persuasion as a source of electoral change.

2 The Private and Public Effects of Campaign Persuasion

This section develops our argument about how campaign appeals affect private preferences and public expression, and why the implications depend on the social environment. We explain how different patterns of effects arise and when they widen or narrow the gap between private preferences and public expression, then turn to the considerations that motivated our predictions for the ad studied here.

We distinguish the preferences people hold privately from the support they express publicly. An

individual may privately support a candidate yet be unwilling to express that support to others, or may express support publicly without privately preferring that candidate. Here we consider public expression to ingroup members, an audience whose approval people value and who are especially likely to sanction deviation from group norms (Hogg and Reid, 2006; Connors, 2020). The ingroup can be defined by gender (Huddy et al., 1997; Klar, 2013), partisanship (Suhay, 2015), or another shared identity.

A campaign appeal may change whether an individual privately supports a candidate and whether they express that support publicly.[3] A person's private support for a candidate depends on the personal value of supporting that candidate weighed against the nonsocial costs, such as the psychological discomfort of departing from one's usual partisan choices even when that choice is not observed by others. Public support depends on these same considerations *and* on social incentives: whether expressing that support to others brings social costs or social benefits. When social norms convey disapproval, publicly expressed support may lag behind privately held support. When they convey approval, publicly expressed support may exceed privately held support, so that people may express support they do not privately hold. A campaign appeal can therefore have different effects on private preferences and public expression because the two outcomes depend on overlapping but distinct considerations.

Campaign appeals can affect private preferences and public expression through two pathways. First, appeals can change *personal beliefs*, altering the personal value of supporting a candidate (Druckman, 2022). They can do so by providing information that changes candidate evaluations (Lodge, Steenbergen and Brau, 1995), increasing the salience of particular issues (Chong and Druckman, 2007), eliciting emotions (Brader, 2006), or activating identities (Klar, 2013; Holman,

[3]For simplicity, we consider binary decisions: whether to support a candidate privately and whether to express support publicly. For empirical purposes, the binary logic can be extended to ordered and continuous outcomes by treating each successively higher level of support as a separate decision about whether to support the candidate at least that much privately or express at least that much support publicly. Changes in the level of support then reflect changes in these decisions. At the population level, effects aggregate changes across individuals, so even binary individual decisions can yield continuously varying average effects.

Schneider and Pondel, 2015). Because personal value shapes both private preferences and public expression, such changes can affect either or both outcomes.

Second, appeals can change *perceived social norms*. Campaign appeals often convey social signals by depicting ingroup members supporting or opposing a candidate.[4] By shifting perceived descriptive or injunctive norms, these appeals can lower the social costs or increase the social benefits of public expression (Cialdini, Reno and Kallgren, 1990; Bursztyn, Egorov and Fiorin, 2020). They can therefore change public expression without changing the personal value of supporting a candidate and thus need not change private preferences.[5] An appeal that lowers perceived social costs without eliminating them makes public expression less costly, but expressing support publicly still requires enough personal value to outweigh both nonsocial and remaining social costs. When public expression instead brings net social benefits, those benefits can lead people to express support they would not hold privately.

The argument thus far builds on established accounts of why private preferences and public expression differ and how campaign appeals can change personal beliefs and perceived norms. Our contribution is to show how changes in personal beliefs and social incentives produce different public–private effect patterns and how these patterns reshape the relationship between private preferences and public expression. We assume that any changes in personal beliefs or social incentives operate in the direction encouraged by the message, although either pathway may remain unchanged.[6] This logic implies four possible patterns: similar effects on both outcomes, no effect

[4]For example, the Trump campaign released a video in 2024 featuring more than a dozen Black women declaring "I'm not with her" (NBC News, 2024). By depicting Black women publicly opposing Harris, the video could challenge the perception that members of this group were expected to support the Democratic candidate.

[5]We do not assume that changes in perceived norms are internalized. Norm shifts may change the personal value people place on supporting a candidate, but they may also produce compliance, in which people publicly conform to social norms they do not personally support (Blinder, Ford and Ivarsflaten, 2013; Willer, Kuwabara and Macy, 2009; Pickup, Kimbrough and de Rooij, 2022).

[6]In our empirical context, this would be increased support for Harris and increased opposition to Trump.

on either, a larger effect on private preferences than on public expression, or a larger effect on public expression than on private preferences. We refer to the first two as *symmetric effects* and the latter two as *asymmetric effects*.

An appeal encouraging support for a candidate can have symmetric effects when individuals initially support the candidate neither privately nor publicly and the message changes their personal beliefs enough—alone or alongside a change in perceived norms—to lead them to support the candidate in both domains. Such effects indicate that the appeal can advance electoral change through both private preferences and public expression, without being confined to either domain. Alternatively, an appeal may have symmetric null effects, leaving both outcomes unchanged. This could occur when individuals already support the candidate privately and publicly, or when the appeal is insufficiently strong to change support in either domain.

At the individual level, asymmetric effects arise when an appeal changes support in one domain but not the other.[7] The asymmetric pattern of central interest here is a larger effect on private preferences than on public expression. This pattern has different implications depending on the social environment. Consider first individuals who face social costs for expressing support and support a candidate neither privately nor publicly. An appeal that increases the personal value of supporting the candidate may lead them to support the candidate privately without expressing that support publicly—even if it also reduces social costs—provided the remaining costs are sufficient to discourage public support. In this way, campaign persuasion can create new gaps between private preferences and public expression and, in the aggregate, deepen preference falsification (Kuran, 1995).

Under social benefits, by contrast, publicly expressed support may initially exceed private support: people express support for the candidate because doing so is rewarded, even though they do not privately support the candidate. A message that raises the personal value of supporting the candidate can shift private preferences toward the position people already express publicly, while public expression remains unchanged. Persuasion then narrows the gap, a pattern consistent with movement from norm compliance towards norm internalization (Kelman, 1958, 2006; Wood, 2000;

[7]Because each support decision is binary, a message cannot increase support in a domain where an individual already supports the candidate.

Willer, Kuwabara and Macy, 2009; Blinder, Ford and Ivarsflaten, 2013; White and Laird, 2020). A central takeaway is that the same pattern of effects—larger effects on private preferences than on public expression—can widen a public–private gap under social costs or narrow one under social benefits.

An asymmetry in the other direction is also possible, with a message having a larger effect on public expression than on private preferences. Under social costs, private support may initially exceed public expression, consistent with pre-existing preference falsification (Kuran, 1995). By increasing the personal value of supporting the candidate or reducing social costs, an appeal can lead individuals to express support they already hold privately, changing public expression while leaving private support unchanged. Under social benefits, individuals who initially support the candidate neither privately nor publicly may come to express support only publicly: an appeal may make support worthwhile when socially rewarded but not in private, either by raising its personal value or by increasing those rewards. Larger effects on public than on private support can thus indicate a narrowing gap when social costs prevail or a widening gap when social benefits prevail.

Overall, the framework yields two implications. First, campaign appeals can affect private preferences and public expression symmetrically or asymmetrically, depending on the intervention pathway, the social environment, and the initial public–private relationship. Second, the social environment and the initial public–private relationship help explain whether a given asymmetry brings private preferences and public expression closer together or moves them further apart. Our preregistered hypotheses focused on larger private effects under social costs, potentially widening the gap. Although we did not anticipate the social benefits interpretation, the framework also explains how larger private effects could narrow a gap where public expression already exceeds private support. The next section describes the ad and electoral context that motivated our predictions.

3 The Campaign Ad

We examine the effects of campaign persuasion on private preferences and public expression using a real-world campaign ad from the 2024 U.S. presidential election in a preregistered study (see Appendix K). The ad, narrated by Julia Roberts, aired during the final week of the campaign amid widespread speculation that some Republican and Independent women disillusioned with Trump might quietly defect to Harris (ABC News, 2024). Reactions to the ad across the political spectrum

suggests it was perceived as especially resonant, even in an already saturated media environment (Newsweek, 2024; Washington Post, 2024).

The ad contained messages that could plausibly shift personal beliefs about the candidates. It emphasized the personal stakes of the election and encouraged women to make their own choice, potentially reducing psychological barriers to privately defecting from Trump among women conflicted about their vote choice. The persuasive potential of these appeals among Republican and Independent women was nevertheless uncertain. Gender-targeted appeals can activate gender identity and increase support among women voters (Holman, Schneider and Pondel, 2015), but partisan identity often outweighs gender identity in shaping political behavior (Huddy, Cassese and Lizotte, 2008). Such appeals can also activate countervailing partisan identities (Klar, 2018), and women in the United States rarely cross party lines on the basis of shared gender identity alone (Junn and Masuoka, 2024). Whether this ad could overcome these barriers only days before the election was therefore an open empirical question.

The ad also conveyed a social signal that could affect public expression. Although it reassured women that dissenting vote choices could remain hidden from their husbands, it depicted women signaling their opposition to Trump to one another. This suggested that opposition among other women was neither isolated nor socially unacceptable. The ad could therefore have reduced the perceived costs or increased the perceived benefits of expressing opposition to Trump to other women, potentially changing public expression independently of any change in private preferences. The "shy voter" account nevertheless led us to anticipate that Republican women in particular might face social costs substantial enough to constrain public expression even if the ad reduced them, motivating our prediction of larger effects in private than in public.

3.1 The Experiment

We test the ad's effects by embedding it in an online survey experiment. Respondents assigned to treatment watched the 30-second ad and answered several reaction questions before proceeding to the post-treatment survey. Respondents assigned to the no-ad control condition proceeded directly

to the post-treatment survey without viewing the ad or answering the reaction questions.[8] Among respondents assigned to treatment, 88 percent reported that they had not previously seen the ad, and 72 percent correctly answered a post-treatment attention check (Appendix C.2).

Using an actual campaign ad trades experimental control over specific features—such as its message, narrator, visuals, and tone—for ecological validity, an approach common in research on political advertising (Clinton and Lapinski, 2004; Coppock, Hill and Vavreck, 2020; Gerber et al., 2011; Huber and Arceneaux, 2007; Hewitt et al., 2024). Our goal is not to isolate which component of the ad produced persuasion, but to investigate whether the ad had differential effects on private versus public expression.

4 Research Design

As detailed in Appendix A, we fielded the online survey with the campaign ad experiment during the final days of the 2024 U.S. presidential election. Eligible participants were voting-age U.S. citizens who identified as women and had not yet voted. Recruitment initially focused on Republican women and was subsequently expanded to Independent and weak or leaning Democratic women.[9] The resulting sample included 2,056 respondents: 37 percent identified as strong Republicans, 31 percent as weak or leaning Republicans, 27 percent as Independents, and 5 percent as weak or leaning Democrats.

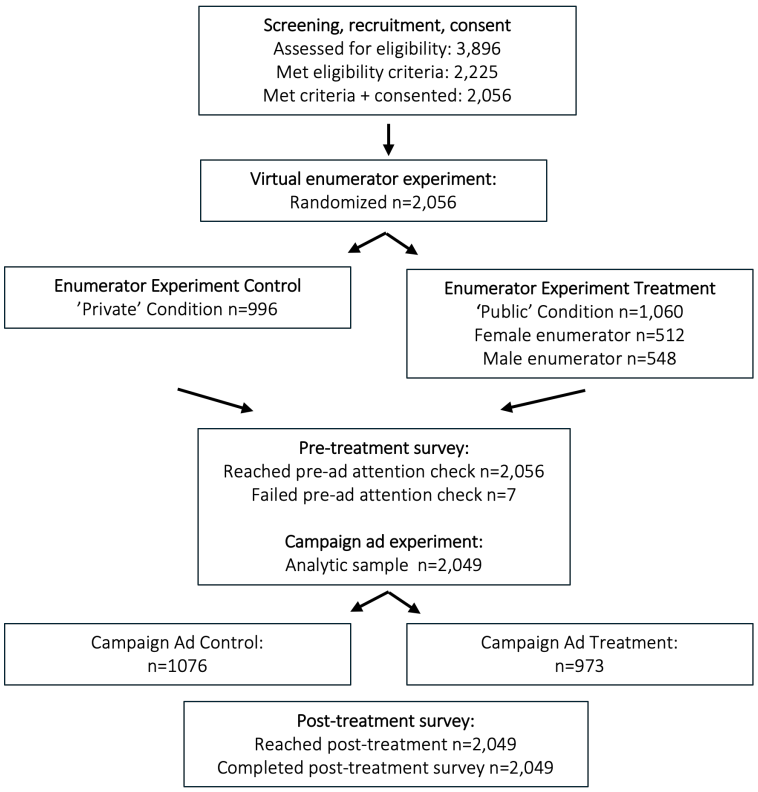
Figure 1 summarizes the overall research design. Following recruitment and consent, participants were randomly assigned to the private or public condition of the virtual enumerator experiment, described below. They then completed a brief pre-ad survey, followed by the campaign ad

[8]The ad had already been publicly released and widely discussed in the media at the time of the study (Newsweek, 2024; Washington Post, 2024). Given the study's relatively small sample and nationwide recruitment strategy, participation was unlikely to have influenced electoral outcomes (Slough, 2024).

[9]The study was approved by the authors' institutional review board (Appendix J) and preregistered. The pre-analysis plan was updated to reflect the change to the sample. The anonymized preregistration is available here; Appendix K documents correspondence between the manuscript and the preregistration plan.

experiment. In the post-ad survey, respondents in the public condition again answered questions posed by a virtual enumerator, while those in the private condition completed the online survey on their own. The survey ended with a measure of private campaign donation behavior. All 2,056 participants who began the study were randomized in the campaign ad experiment. Following the pre-analysis plan, we excluded seven who failed the pre-ad attention check, yielding an analytic sample of 2,049.[10]

Figure 1: Research Design Overview



4.1 The Virtual Enumerator Experiment

The virtual enumerator experiment serves two purposes: to examine differences between private preferences and public expression before ad exposure, and to compare the ad's effects on these outcomes while holding all other features of the attitude measures constant. Before the pre-ad

[10]Balance checks show no statistically significant joint imbalance between the ad and control or female-enumerator and private conditions, although some baseline characteristics differed between the male-enumerator and private conditions (Appendix C.1).

survey, respondents were randomly assigned to a *private* condition, in which they answered a standard self-administered online survey, or a *public* condition, in which they answered the same questions posed by a female or male virtual co-partisan enumerator. Varying the enumerator's gender allows us to examine whether expression and ad effects differ between female and male audiences, as discussed below. Respondents in the public condition were further randomized among six enumerators, three women and three men, so that the manipulation was not confounded with the characteristics of any single individual (see Appendix A.2). The enumerator was introduced through an image and a brief chat-style exchange, and the image remained visible throughout the survey. Across conditions, respondents answered the same direct questions, with identical response options and formats; the virtual enumerator was present only in the public condition, throughout both the pre-ad and post-ad surveys.[11]

The private condition provides the closest observable approximation of respondents' explicitly reported private preferences. This interpretation is consistent with evidence that self-administered surveys are perceived as more private and tend to elicit more candid answers to sensitive questions than interviewer-administered modes (Tourangeau and Yan, 2007; Holbrook and Krosnick, 2010; Krumpal, 2013; Bursztyn et al., 2025). It does not require assuming that anonymous self-administration perfectly reveals respondents' true beliefs. Respondents may anticipate broader audiences or otherwise adapt their answers even when completing surveys privately (Lindsey, 2023; Krupnikov, Piston and Bauer, 2016). The relevant claim is comparative: anonymous self-administration more closely approximates private expression than reporting the same views in the presence of a socially identified audience.

The public condition introduces a socially relevant audience for political expression. A large literature shows that respondents adjust survey answers to perceived audiences and their expectations, and that interviewer characteristics such as gender, race, and ethnicity can affect the views respondents express (Tourangeau and Yan, 2007; Huddy et al., 1997; Adida et al., 2016; Blair, Coppock and Moor, 2020; Bursztyn et al., 2025). Interviewers can serve as salient social referents

[11]Because the enumerator was present before ad exposure, audience assignment could also have influenced message processing. Interpreting differences in ad effects solely as differences in expression assumes that the audience did not differentially change respondents' underlying response to the ad.

when respondents infer that they hold expectations about which attitudes are appropriate to express (Blair, Coppock and Moor, 2020). Responses may therefore reflect perceived norms and the anticipated costs or benefits of expressing particular views.[12] Research also finds that respondents react to virtual interviewers in socially meaningful ways (Krysan and Couper, 2003).

We thus interpret the virtual enumerator as introducing a perceived audience for social evaluation into an otherwise anonymous survey. Specifically, the enumerator captures expression to a general ingroup member: an unfamiliar person whose social relevance derives from shared group membership—partisanship and, for female enumerators, gender—rather than from a personal relationship. This approach has two advantages for our purposes. First, because respondents have no history with the enumerator, their responses can reflect perceived norms of the shared group rather than the views or anticipated reactions of particular friends or family members, which vary across respondents and cannot be randomly assigned.[13] Second, a general ingroup member is the audience the ad's social signal addresses: the ad depicts women signaling their choices to other women, not to particular members of a viewer's network. In the conclusion, we discuss the extent to which findings from this design generalize to public expression within respondents' own social networks.

We identified all virtual enumerators as co-partisans to reduce ambiguity about the audience's partisan affiliation. Without this information, some respondents might have perceived the enumerator as a co-partisan and others as a non-co-partisan, introducing uncontrolled variation in the social incentives surrounding expression. Second, doing so provided a substantively relevant setting for political expression, given that political discussion often occurs within relatively homogeneous partisan networks (McPherson, Smith-Lovin and Cook, 2001; Mutz, 2006). Shared partisanship did not necessarily define the audience's relevance, however: respondents could perceive the enumerator as an ingroup member on the basis of partisanship, gender, or both.

[12]The design does not distinguish among anticipated social evaluation, activation of partisan or gender identity, and heightened awareness of group norms. We therefore interpret differences across conditions broadly as audience effects rather than attributing them to a single psychological process.

[13]Manipulating expression to respondents' actual network members would also have required disclosing their survey responses to friends or family, posing substantial ethical and practical barriers.

Our primary comparison is between the private and female enumerator conditions because the ad's portrayal of women signaling their choices to one another should most directly affect political expression to another woman. At the same time, the portrayal of women concealing their choices from their husbands could reinforce perceived costs of expressing opposition to Trump to men, further motivating the comparison between female and male audiences. The male enumerator condition allows us to examine whether expression and ad effects vary with audience gender while holding stated partisan affiliation constant. Because both female and male virtual enumerators introduced a socially identified audience and disclosed a shared partisan affiliation, differences between them are less readily attributable to the mere presence of a virtual enumerator or to a general reminder of partisan identity.

Treatment checks indicate that respondents registered the manipulation, although they identified the enumerator's gender more accurately than partisan affiliation (Appendix C.2). This supports interpreting gender as a salient audience cue, while leaving open the extent to which shared partisanship also shaped responses. We therefore describe the public conditions as expression to female or male enumerators identified as co-partisans, without assuming that respondents attached equal weight to both identities.

4.2 Outcome Measures, Hypotheses, and Estimation

We examine the effect of the ad on private and publicly expressed support for Trump and Harris. Our primary outcomes are the preregistered candidate support indices, which combine candidate favorability and intended vote choice. We also report the individual components to clarify the patterns underlying the index effects. A behavioral measure of willingness to make a private donation provides a secondary outcome, which we discuss where relevant to the interpretation of the main findings. Details on outcome measurement and index construction are provided in Appendix D.

Overall, the design allows us to estimate the ad's effects on preferences as privately held versus publicly expressed to a virtual enumerator. Following our pre-registered predictions (see Appendix E.1), we assess the following main hypotheses:

- The ad decreases private support for Trump.

- The ad increases private support for Harris.
- The ad decreases Trump support more in private than in public.
- The ad increases Harris support more in private than in public.

The key quantity for the latter two predictions is the public-minus-private difference in ad effects. In our main analysis below, we define public expression as expression to the female virtual enumerator because the ad depicts women signaling their choices to one another. The pre-analysis plan pooled expression to female and male enumerators and separately predicted larger ad effects on expression to female than to male enumerators, but did not explicitly specify the female-public versus private hypotheses (see Appendix K). Our focus on these contrasts follows the substantive reasoning behind those predictions. They are evaluated using coefficients in the prespecified regression summarized below, requiring no change to the model or estimation procedure, but remain distinct from the explicitly preregistered pooled-public comparisons, which we also report.

We intentionally did not preregister a separate prediction for the public effect alone because the predictions for the private effect and the public–private difference already capture the asymmetry of interest. The public effect may be null or move in the intended direction, provided that it is smaller in magnitude than the private effect. Although these predictions reflect our preregistered expectations in the "shy voter" context, the design allows us to assess any of the public–private effect patterns described in the theory section. Appendix E.1 details our preregistered expectations about baseline public–private differences and heterogeneity by respondent party.

As detailed in Appendix E.2, we estimate the campaign ad's effects on privately held and publicly expressed preferences using the following pre-registered specification:

$$\begin{aligned} Y_i = \alpha &+ \beta_1 Ad_i + \beta_2(Ad_i \times PubFem_i) + \beta_3 PubFem_i \\ &+ \beta_4(Ad_i \times PubMale_i) + \beta_5 PubMale_i + X_i'\psi + \mu_{d[i]} + \epsilon_i. \end{aligned}$$

where Y_i is the outcome for respondent i; Ad_i indicates assignment to the campaign ad; and $PubFem_i$ and $PubMale_i$ indicate assignment to the female or male virtual enumerator, respectively.

The private condition is the reference category. β_1 estimates the ad's effect in the private condition, and β_2 estimates the public-minus-private difference in ad effects for expression to the female enumerator. For Harris support, we predict a positive private effect ($\beta_1 > 0$) and a negative

public-minus-private difference ($\beta_2 < 0$), indicating a smaller increase in support in public than in private. For Trump support, we predict a negative private effect ($\beta_1 < 0$) and a positive public-minus-private difference ($\beta_2 > 0$), indicating a smaller decrease in support in public than in private.[14] We also report the ad's effect in the female-public condition, $\beta_1 + \beta_2$, for completeness. β_3 estimates the effect of expression to the female enumerator relative to private expression among respondents in the no-ad control condition, which we use to assess social incentives in the absence of the ad. All specifications include the prespecified demographic controls X_i and fixed effects $\mu_{d[i]}$ for the day on which respondent i completed the study, and use heteroskedasticity-robust standard errors.

To examine whether these effects vary by respondent party, we estimate a fully interacted specification that includes the triple interaction of the ad indicator, the enumerator indicators, and party identification. This allows both the private ad effect and the public-minus-private difference to vary between Republican and Independent women. We calculate the party-specific effects as linear combinations of the relevant coefficients, and the triple interaction coefficient tests whether the public-minus-private difference differs between the two groups. We report party specific effects in the main text and respective appendices.

5 Descriptive Statistics

Appendix B presents descriptive statistics using pre-ad data from respondents in the private condition. These suggest potential scope for persuasion even in the final days of the campaign. Overall, 68 percent were very certain of their vote choice, leaving nearly one-third less certain. Certainty varied sharply by party: 81 percent of Republican women were very certain, compared with 38 percent of Independent women.[15] Similarly, 83 percent of Republican women said they were very likely to vote for Trump, compared with about 25 percent of Independent women. Independent

[14]Because these hypotheses make directional predictions, the pre-analysis plan specifies one-sided tests. We nonetheless report conventional two-sided p-values throughout, which are more conservative for estimates in the predicted direction. Multiple hypothesis correction tests in Appendix I report corrections using both one-sided and two-sided p-values.

[15]We do not discuss weak/leaning Democrats here due to their small sample size.

women thus entered the experiment with less settled vote choices and substantially weaker Trump support.

Republican and Independent women also described different social environments. Republican women's networks were predominantly Trump-supporting: 77 percent said most of their friends and family were voting for Trump, and among those with a partner, 89 percent said their partner would vote for Trump. Independent women's networks were more mixed: the corresponding figures were 34 and 41 percent. These differences suggest that the social incentives surrounding public expression could vary by party. Republican women's predominantly Trump-supporting networks could make expressing opposition socially costly, whereas Independent women's more mixed networks leave open whether such expression carried costs or benefits.

6 Ad Effects on Private and Public Candidate Support

Figure 2 presents the main public–private campaign ad treatment effects, showing effects in the private condition (top), public (female enumerator) condition (middle), and the public–private difference (bottom). Appendix Table G.2.1 reports the full regression results.

The results show clear evidence of the predicted asymmetry for Trump. The campaign ad reduced private Trump support by .10 SD on the index ($p = .044$). Both components declined, with favorability falling by .16 units ($p = .034$) and intended vote by four percentage points ($p = .102$). However, these effects did not appear in the public (female enumerator) condition: the ad had no detectable effect on publicly expressed Trump support on the index (.07 SD, $p = .318$) or either component. The public–private difference in the index effect was therefore .17 SD ($p = .049$), consistent with the expectation that the ad would cause a bigger reduction in private relative to public support.[16] The explicitly preregistered private–public contrasts pooling female and male enumerators show a similar pattern, with slightly smaller differences in ad effects (Appendix G.1).

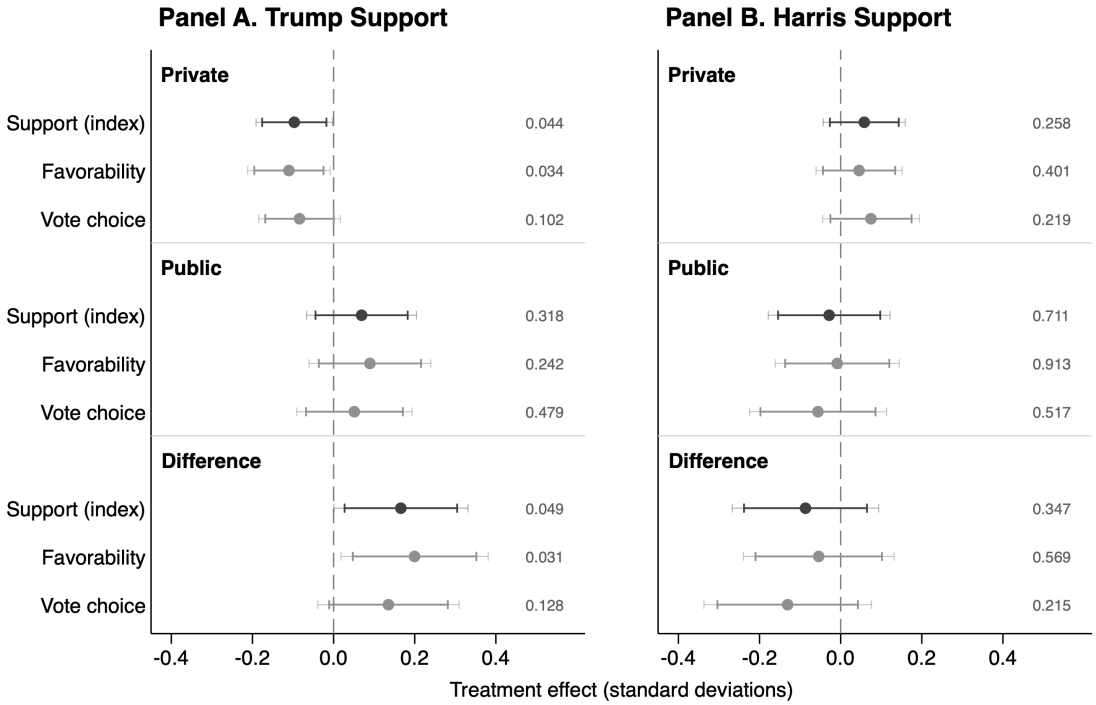
The Harris results suggest a mirror-image pattern, though estimates are less precise. In the private condition, the index effect is positive but not statistically significant (.06 SD, $p = .258$). One potential explanation for this null result is that the ad shifted private support towards third-party

[16]Appendix I reports sharpened false discovery rate q-values for the primary persuasion tests and exploratory supporting analyses.

candidates: results in Appendix Table G.2.1 show the ad increased the likelihood of voting for a third-party candidate by three percentage points ($p = .019$). However, similar to the Trump results, the ad had no detectable effect on publicly expressed support for Harris or likelihood of voting for a third party. The coefficient on the public–private difference in the Harris index is correspondingly negative—reflecting the expected smaller increase in public relative to private Harris support—but not statistically significant ($-.09$ SD, $p = .347$); its components follow the same pattern.

Overall, these results suggest that the campaign ad affected privately held and publicly expressed attitudes asymmetrically in the full sample, most clearly for Trump support. Although the pattern is less clear for Harris—in part because shifts in private support away from Trump appear to have benefited third-party candidates as well—it nonetheless demonstrates that the same campaign message can affect private preferences and public expression differently. We turn next to examining what accounts for this asymmetry.

Figure 2: Ad Effects on Private and Public Candidate Support



Notes: Points are estimated treatment effects of exposure to the ad based on regressions with fixed effects, controls, and heteroskedasticity-robust standard errors. All outcomes are standardized so their effects are in standard deviation units. Private estimates are treatment effects among respondents in the private condition. Public estimates are treatment effects among respondents who answered to a female virtual enumerator. The difference rows show the public effect minus the private effect, estimated as the interaction between treatment and the female-enumerator condition. Dark circles indicate the main support index, which combines the favorability and vote choice components (light circles). Thick bars show 90 percent confidence intervals and thin bars show 95 percent confidence intervals. Numbers at right are two-sided p-values.

7 Explaining Ad Effects on Private but not Public Support

Why did the campaign ad change private but not public Trump support? Our preregistered expectation attributed such an asymmetry to social costs for Republican women: the ad would shift private preferences, but even if it shifted norms and reduced those costs, the remaining costs would deter public opposition to Trump.

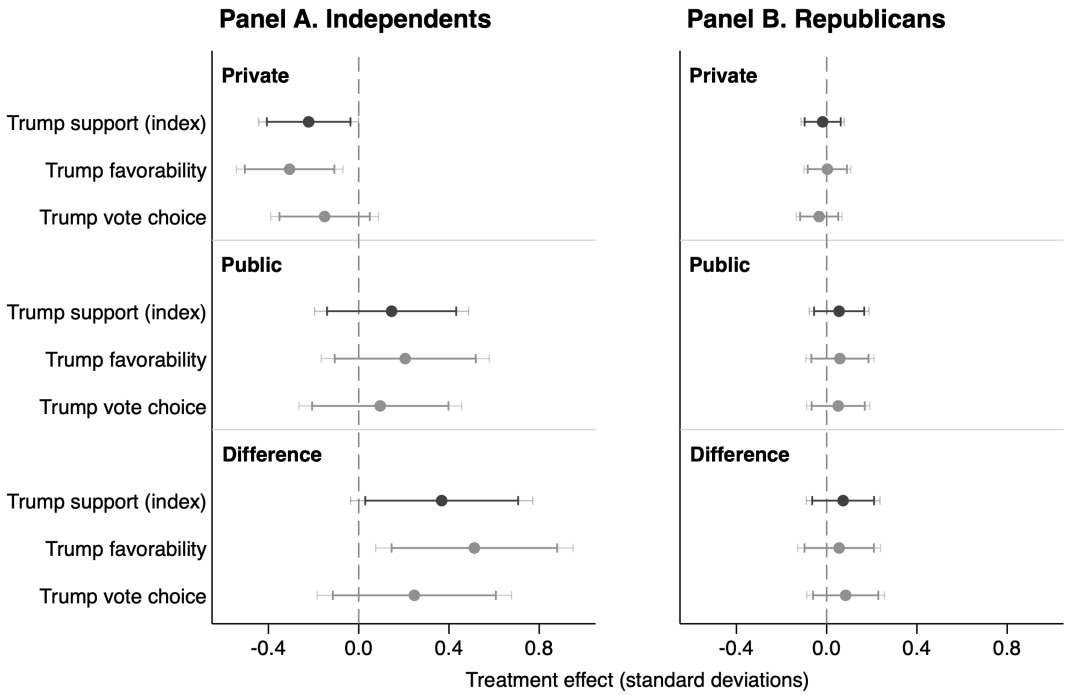
We begin by disaggregating the main results by respondent party.[17] Figure 3 presents the party-specific Trump results; detailed regression estimates and the corresponding Harris results appear in Appendix G.3. As can be seen in Panel B, the ad had no detectable effect on either private or public Trump support among Republican women. We therefore find no evidence of the predicted "shy voters" pattern. Instead, the asymmetric effects were most clearly evident among Independent women. The ad reduced their private Trump support index by .22 standard deviations ($p = .049$), reflecting declines in both favorability and vote choice. It also reduced Independent women's likelihood of clicking through to Trump's donation page in the private condition by 3 percentage points ($p = .033$; Appendix Table G.3.1), indicating that the shift in privately held attitudes extended to private behavior.

Yet the ad had no detectable effect on Independents' publicly expressed Trump support. The resulting public–private difference was .37 SD ($p = .074$), indicating a larger decline in private relative to public support. The difference in this contrast between Independent and Republican women is not statistically significant (.24 SD, $p = .284$), so we do not claim that the two groups differ (see Appendix Table G.3.2). Our claim is narrower: the pooled pattern is most clearly evident among Independent women. Their private response is consistent with evidence that gender-based appeals in campaign advertisements can persuade Independents (Kam, Archer and Geer, 2017). It leaves open, however, why their public expression did not change.

One possibility is that Independent women, like the shy voters we anticipated among Republican women, faced social costs that kept them from expressing a change in private preferences. The descriptive statistics give some reason for doubt: Independent women described more politically

[17] We preregistered examining heterogeneous effects by party as an exploratory analysis (see Appendix K). We exclude weak and leaning Democrats because of their small sample size.

Figure 3: Ad Effects on Private and Public Trump Support by Party



Notes: Points are estimated treatment effects of exposure to the ad based on regressions with fixed effects, controls, and heteroskedasticity-robust standard errors, estimated separately for Independents (Panel A, $n = 555$) and Republicans (Panel B, $n = 1399$). All items are standardized so their effects are in standard deviation units. Private estimates are treatment effects among respondents in the private condition. Public estimates are treatment effects among respondents who answered to a female virtual enumerator. The difference rows show the public effect minus the private effect, estimated as the interaction between treatment and the female-enumerator condition. Dark circles indicate the main support index, which combines the favorability and vote choice components (light circles). Thick bars show 90 percent confidence intervals and thin bars show 95 percent confidence intervals.

mixed social environments than Republican women, suggesting that opposing Trump was less likely to carry social costs.

Figure 4 provides a more direct test. It plots mean post-ad Trump support in the ad and control groups, both as privately held and as publicly expressed to the copartisan female enumerator. If women faced social costs for opposing Trump, we would expect two features. First, without the ad, publicly expressed Trump support should match or exceed privately held support: social costs could lead some women to express more support for Trump than they privately held, but would give them no reason to express less. Second, if the ad reduced private Trump support while social costs remained, women who came to oppose Trump would keep that opposition to themselves, so public support would stay near its level without the ad while private support fell below it. We expected this pattern primarily among Republican women, for whom the prevailing ingroup norm favors Trump. As the bottom panel shows, however, their Trump support was similar across all

four conditions, so the ad produced no private shift for social costs to keep from public view.

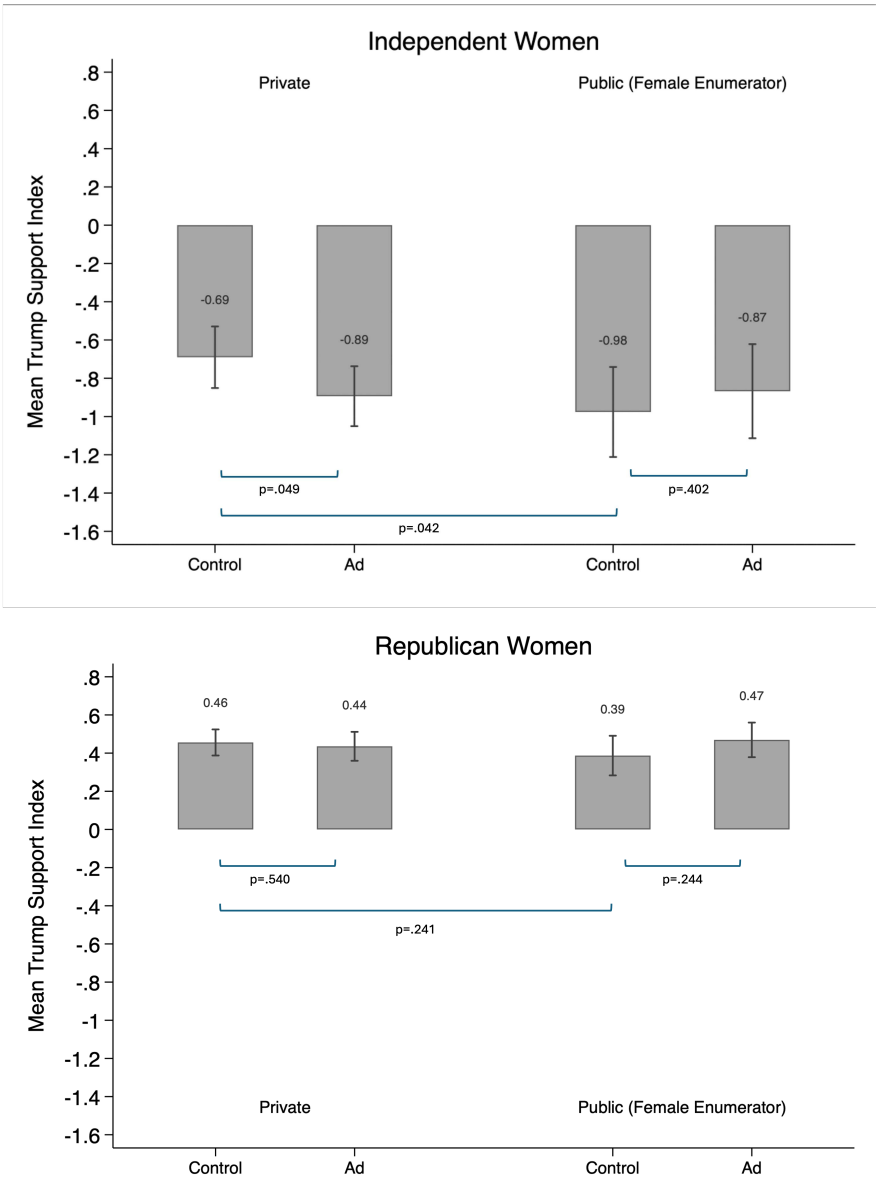
The top panel shows that Independent women departed from both expectations. Among these women, privately held Trump support was lower in the ad condition than in the control condition ($-.89$ versus $-.69$), consistent with the regression-based treatment effect reported above. Publicly expressed support, however, was already low in the absence of the ad: in the control condition, Independent women expressed less support for Trump to a female copartisan than they held privately ($-.98$ versus $-.69$, $p = .042$; Appendix Table G.3.1). Relative to this public control level, the ad had no additional detectable effect on publicly expressed support ($-.87$ versus $-.98$, $p = .402$). Following exposure, privately held and publicly expressed support were at similar levels ($-.89$ and $-.87$). This pattern does not fit the shy voter account in which social costs prevent women from publicly expressing an ad-induced private change: under that account, publicly expressed support would remain above privately held support. Instead, the ad reduced privately held support toward the level Independent women already expressed publicly.

7.1 The Effects of Public Expression (Pre-Ad)

These patterns are consistent with social benefits, rather than costs, to expressing opposition to Trump to a female copartisan, but they rest on comparisons within the post-ad experiment. Additional evidence comes from responses collected before respondents saw any campaign ad. Because these outcomes were measured before ad exposure, any difference between public and private responses reflects the audience alone. If Independent women perceived social benefits to expressing opposition to Trump to another Independent woman, then expressing their views to the co-partisan female enumerator should reduce Trump support relative to private expression. That effect should also be specific to a female audience, rather than appearing for any co-partisan enumerator. We therefore estimate the effects of public expression before the ad experiment, comparing responses to the female and male enumerators with private expression. Figure 5 presents these effects, and Appendix Table H.1.1 reports the corresponding regression results.

The figure shows that reporting to the female copartisan enumerator shifted Independent women's responses in a direction consistent with social incentives to express opposition to Trump and concern for reproductive rights. Relative to the private condition, expression to the female virtual enumerator reduced their reported likelihood of voting for Trump by .38 points on a 1-5

Figure 4: Mean Levels of Trump Support by Ad and Public Female Enumerator Treatment Conditions

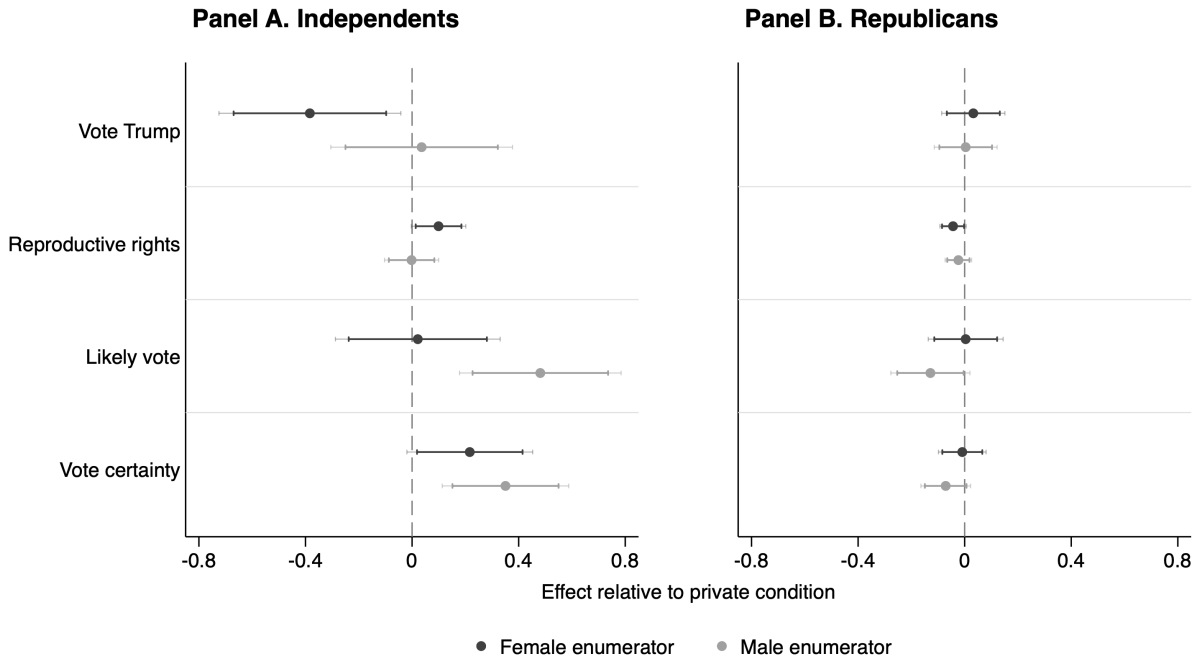


Note: Bars report raw mean values of the Trump support index by treatment condition separately for Independent and Republican women. The private condition measures preferences reported directly in the survey, while the public condition measures preferences reported to a female virtual enumerator. Error bars represent 95 percent confidence intervals. Reported p-values are from regression models corresponding to Appendix Table G.3.1

scale ($p = .028$). The pattern extended to issue priorities: reporting to the co-partisan female enumerator increased the probability that Independent women named reproductive rights as one of the country's three most important issues by 10 percentage points ($p = .057$). These pre-ad results closely resemble the public expression patterns among Independent women in the no-ad condition.

These shifts depended on the gender of the audience. Expression to the male copartisan enumer-

Figure 5: Effects of Public Expression to Female and Male Enumerators, by Party



Notes: Points are estimated effects of expression to a virtual enumerator based on regressions with fixed effects, controls, and heteroskedasticity-robust standard errors, estimated separately for Independents (Panel A) and Republicans (Panel B). Controls exclude party identification. All effects are relative to the private condition and are in the original units of each outcome. Female enumerator estimates are effects among respondents who answered to a female virtual enumerator, and male enumerator estimates are effects among respondents who answered to a male virtual enumerator. Estimates pool respondents across the ad treatment and control conditions. Thick bars show 90 percent confidence intervals and thin bars show 95 percent confidence intervals.

ator had no detectable effect on Independent women's likelihood of voting for Trump (.04 points, $p = .836$), and the female–male difference was statistically significant ($-.42$ points, $p = .045$). The male enumerator likewise had no detectable effect on naming reproductive rights as a top issue (0 percentage points, $p = .972$), although the female–male difference was less precisely estimated (10 percentage points, $p = .100$).

The absence of these shifts with the male enumerator did not reflect a general lack of responsiveness to that audience. Relative to private expression, expression to the male co-partisan enumerator increased Independent women's expressed likelihood of voting and certainty about their vote choice. The male audience thus affected expressions of general political engagement, while lower Trump support and greater concern about reproductive rights appeared only with the female audience. Nor do we find evidence that social costs of expressing opposition to Trump or concern about reproductive rights to the male audience account for the female–male difference: relative to private

expression, the male enumerator produced no detectable increase in expressed Trump support or decrease in concern about reproductive rights. The contrast instead reflects lower Trump support and greater concern about reproductive rights with the female audience, consistent with social benefits to expressing these views to another woman identified as an Independent.

These results also challenge an account in which a general reluctance to appear partisan operates similarly across audiences. Many self-identified Independents hold partisan preferences but mask them in social settings because partisanship is perceived as socially undesirable (Klar and Krupnikov, 2016). Independent women might therefore understate their Trump support to any audience simply to avoid appearing partisan. That account predicts lower expressed Trump support with both enumerators. Instead, the shift appeared only with the female enumerator, indicating patterns are inconsistent with a general reluctance to appear partisan.

We do not see the same pattern among Republican women. Expression to the co-partisan female enumerator had no detectable effect on their expressed likelihood of voting for Trump (.03 points, $p = .591$) and, if anything, reduced the probability that they named reproductive rights among their top issues by 4 percentage points ($p = .083$). The female-enumerator effects differed between Independent and Republican women by $-.42$ points for likelihood of voting for Trump ($p = .020$) and by 14 percentage points for naming reproductive rights ($p = .017$). Although between-party differences are not central to our argument, these results suggest that the shifts among Independent women reflected the particular co-partisan audience they faced, rather than a general response to expressing political views to another woman.

7.2 Evidence for Social Norm Change

We also consider whether the absence of an ad effect on public expression reflects a failure to change perceived ingroup norms. The ad's social signal was expected to shift perceptions of prevailing norms among women and, in turn, to reduce the social costs or increase the social benefits of public expression to other women. If the ad did not shift those perceptions, that failure could explain its null effect on publicly expressed Trump support. Appendix H.2 examines the ad's effects on respondents' perceived descriptive norms, measured as the belief that more than half of Independent women, Republican women, or Independent men would vote for Harris.

The results suggest that the ad shifted perceived norms. Among Independent women, it in-

creased the proportion expressing privately that more than half of Independent women would vote for Harris by 10 percentage points, significant at the 10 percent level ($p = .054$). The absence of a detectable public effect therefore does not appear to reflect simply a failure to shift perceived norms.

Rather, the evidence shows that publicly expressed perceptions of Harris support were already elevated in the absence of the ad. Among Independent women in the ad control condition, expression to the female copartisan enumerator increased the likelihood of expressing that more than half of Independent women would vote for Harris by 19 percentage points relative to private expression ($p = .005$). This pattern is consistent with social benefits: in telling another Independent woman that most Independent women would vote for Harris, respondents may have been affirming what they took to be their group's norm. The norms results thus parallel those for Trump support: the ad shifted privately expressed perceptions toward an already elevated public level, with no detectable effect on public expression (-13 percentage points, $p = .131$). The resulting public–private difference in ad effects was -23 percentage points ($p = .020$).

The specificity of the estimates provides additional support for this interpretation: the pattern is most evident when Independent women express beliefs about their own group to another Independent woman. We find no evidence of such asymmetric ad effects when they express beliefs about Independent women to a male copartisan enumerator (Appendix Table H.2.1). This asymmetry also differs significantly from the corresponding differences for beliefs about Republican women and Independent men (Appendix H.2). These comparisons support the interpretation that the pattern reflects public expression about an ingroup to an audience sharing both party and gender.

Taken together, the audience effects and perceived-norm results support a social benefits interpretation of the asymmetry among Independent women. Their public opposition to Trump exceeded private opposition without the ad, and exposure brought private preferences closer to that existing public position. This pattern contrasts with the anticipated account in which social costs leave ad-induced private opposition concealed. Appendix H.3 addresses additional alternative explanations.

8 Conclusion

This paper addresses two central questions. Does persuasion change what people privately believe, what they are willing to express publicly, or both? And how do these effects reshape the relationship between private preferences and public expression? We show that a campaign ad reduced privately held support for Trump but had no detectable effect on support expressed to a female copartisan enumerator, producing a significant difference between its private and public effects. The evidence does not support the preregistered expectation that social costs among Republican women prevented private preference change from translating into public expression. Instead, in the absence of the ad, Independent women expressed less support for Trump publicly than privately—a pattern consistent with social benefits of expressing opposition to another Independent woman—and the ad brought their private preferences closer to what they already expressed publicly. These findings show that campaign persuasion can have asymmetric effects on private preferences and public expression. Studying both reveals what a message changes and whether it widens or narrows the gap between the two, providing a fuller account of how campaign persuasion can contribute to electoral change.

Our findings suggest that the relationship between persuasion, private preferences, and public expression varies across groups and contexts. The contrast between Independent and Republican women is instructive. Independent women entered the study with weaker partisan attachments, greater uncertainty about their vote choice, and more politically heterogeneous social environments—conditions that may have increased the scope for persuasion for this group. More broadly, generalizing the public–private effect pattern requires considering whether the social environment rewards or penalizes the position expressed and the initial relationship between private preferences and public expression. Together, these conditions shape which effect patterns are possible and whether persuasion widens or narrows a gap. Our study thus responds to calls for greater attention to the contexts of persuasion (Druckman, 2022) by identifying social incentives and the preexisting public–private relationship as contextual features that help explain what campaign appeals change.

Just as the relationship between private preferences and public expression is likely to vary across groups, it is also likely to vary with the audience. Here, public expression was captured by what

respondents were willing to express to an unfamiliar virtual enumerator. We provide substantial evidence that the female virtual enumerator was perceived as a socially salient audience. Still, these results should be interpreted as public expression shaped by generally perceived ingroup norms rather than as public expression to an actual network member. Generalizing beyond this setting likely requires considering two aspects of social incentives: their direction (whether expressing a position brings social benefits or costs) and their weight (how much those benefits or costs matter to the individual). Our findings suggest that direction mattered and that asymmetric effects emerged even in a setting where relational stakes were limited. Within real-world networks, the direction of social pressure may vary not only with the political views of an individual audience (Carlson and Settle, 2016) but also with the homogeneity or heterogeneity of the network itself (McPherson, Smith-Lovin and Cook, 2001). The weight of social incentives may also vary. It may be limited when individuals are relatively unaffected by others' reactions or when trust makes disagreement less threatening. Conversely, richer relationships, repeated interactions, and greater reputational stakes may give social incentives more weight. Where real-world relationships carry more weight, asymmetries of the kind observed here may be more pronounced, and investigating when this holds is an important task for future research.

Beyond whether these patterns extend to other audiences, their electoral significance depends on whether they translate into behavior and persist over time. Although our outcomes are primarily attitudinal, the ad also reduced clicks to Trump's donation page among Independent women, providing evidence of a behavioral consequence. We do not directly observe vote choice or social influence, but existing research suggests plausible links to both. Because attitudes predict behavior best under corresponding conditions (Ajzen and Fishbein, 1977), private preferences may be especially relevant to behaviors undertaken outside others' view, including vote choice (Kelley and Mirer, 1974; Glasman and Albarracín, 2006), while publicly expressed positions may be more closely related to visible behavior. Public expression is also political behavior in its own right (Makse and Sokhey, 2014): it can shape perceived norms and interpersonal persuasion (Kuran, 1995; Mutz, 2006) and commit people to later action (Pallak, Cook and Sullivan, 1980; Lokhorst et al., 2013).

These consequences also depend on persistence. Persuasion effects often decay quickly (Hill et al., 2013; Gerber et al., 2011), although some interventions produce attitude change that lasts for weeks or months (Broockman and Kalla, 2016; Coppock, Ekins and Kirby, 2018), and private

preferences and public expression may change at different rates (Kelman, 1958; Prentice and Miller, 1993; Kuran, 1995). If the private shift we observe fades while public expression holds steady, the preexisting gap would widen again. Our immediate measures show where the ad initially produces change, not where its effects ultimately settle. Measuring both private preferences and public expression repeatedly over time would help establish how the relationship between the two continues to evolve and where the barriers to change lie. More broadly, our findings show why measuring both outcomes matters: a larger effect on private preferences can narrow, rather than widen, an existing public–private gap. Understanding campaign persuasion therefore requires attention to what changes privately, what changes publicly, and how the social environment shapes the relationship between the two.

Use of Generative AI

Generative AI (ChatGPT 6.1 and Claude Opus 5.5, as well as earlier models) was used to assist with drafting and revising text, organizing and clarifying theoretical arguments, and presenting and interpreting results, including producing figures from regression code supplied by the authors. The authors reviewed and verified all AI-assisted outputs and take responsibility for the manuscript's content.

References

ABC News. 2024. "Gender Gap Could Decide 2024 Election.". Published 2024.

Adida, Claire, Karen Ferree, Daniel Posner and Amanda Robinson. 2016. "Who's Asking? Interviewer Coethnicity Effects in African Survey Data." *Comparative Political Studies* 49(12):1630–1660.

Ajzen, Icek and Martin Fishbein. 1977. "Attitude–Behavior Relations: A Theoretical Analysis and Review of Empirical Research." *Psychological Bulletin* 84(5):888–918.

Blair, Graeme, Alex Coppock and Margaret Moor. 2020. "When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments." *American Political Science Review* .

Blinder, Scott, Robert Ford and Elisabeth Ivarsflaten. 2013. "The better angels of our nature: How the antiprejudice norm affects policy and party preferences in Great Britain and Germany." *American Journal of Political Science* 57(4):841–857.

Bond, Robert M, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle and James H Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization." *Nature* 489(7415):295–298.

Brader, Ted. 2006. *Campaigning for hearts and minds: How emotional appeals in political ads work*. University of Chicago Press.

Broockman, David E and Joshua L Kalla. 2023. "When and why are campaigns' persuasive effects small? Evidence from the 2020 US presidential election." *American Journal of Political Science* 67(4):833–849.

Broockman, David and Joshua Kalla. 2016. "Durably Reducing Transphobia: A Field Experiment on Door-to-Door Canvassing." *Science* 352(6282):220–224.

Bursztyn, Leonardo, Alessandra L González and David Yanagizawa-Drott. 2020. "Misperceived social norms: Women working outside the home in Saudi Arabia." *American economic review* 110(10):2997–3029.

Bursztyn, Leonardo, Georgy Egorov and Stefano Fiorin. 2020. "From extreme to mainstream: The erosion of social norms." *American economic review* 110(11):3522–3548.

Bursztyn, Leonardo, Ingar K Haaland, Nicolas Röver and Christopher Roth. 2025. The social desirability atlas. Technical report National Bureau of Economic Research.

Carlson, Taylor N. 2019. "Through the Grapevine: Informational Consequences of Interpersonal Political Communication." *American Political Science Review* 113(2):325–339.

Carlson, Taylor N. and Jaime E. Settle. 2016. "Political Chameleons: An Exploration of Conformity in Political Discussions." *Political Behavior* 38(4):817–859.

Chong, Dennis and James N Druckman. 2007. "Framing theory." *Annu. Rev. Polit. Sci.* 10(1):103–126.

Cialdini, Robert B, Raymond R Reno and Carl A Kallgren. 1990. "A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places." *Journal of personality and social psychology* 58(6):1015.

Clinton, Joshua D and John S Lapinski. 2004. ""Targeted" advertising and voter turnout: An experimental study of the 2000 presidential election." *The Journal of Politics* 66(1):69–96.

CNN. 2024. "How Liz Cheney tried to give GOP women permission to vote for Harris.". Published October 22, 2024.

Connors, Elizabeth C. 2020. "The Social Dimension of Political Values." *Political Behavior* 42(3):961–982.

Coppock, Alexander, Emily Ekins and David Kirby. 2018. "The Long-Lasting Effects of Newspaper Op-Eds on Public Opinion." *Quarterly Journal of Political Science* 13(1):59–87.

Coppock, Alexander, Seth J Hill and Lynn Vavreck. 2020. "The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments." *Science advances* 6(36):eabc4046.

Druckman, James N. 2022. "A Framework for the Study of Persuasion." *Annual Review of Political Science* 25:65–88.

Druckman, James N, Matthew S Levendusky and Audrey McLain. 2018. "No need to watch: How the effects of partisan media can spread via interpersonal discussions." *American Journal of Political Science* 62(1):99–112.

Gerber, Alan S, Gregory A Huber, David Doherty and Conor M Dowling. 2013. "Is there a secret ballot? Ballot secrecy perceptions and their implications for voting behaviour." *British Journal of Political Science* 43(1):77–102.

Gerber, Alan S., James G. Gimpel, Donald P. Green and Daron R. Shaw. 2011. "How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *American Political Science Review* 105(1):135–150.

Glasman, Laura R. and Dolores Albarracín. 2006. "Forming Attitudes That Predict Future Behavior: A Meta-Analysis of the Attitude–Behavior Relation." *Psychological Bulletin* 132(5):778–822.

Hayes, Andrew F, Dietram A Scheufele and Michael E Huge. 2006. "Nonparticipation as self-censorship: Publicly observable political activity in a polarized opinion climate." *Political Behavior* 28(3):259–283.

Hewitt, Luke, David Broockman, Alexander Coppock, Ben M Tappin, James Slezak, Valerie Coffman, Nathaniel Lubin and Mohammad Hamidian. 2024. "How experiments help campaigns persuade voters: Evidence from a large archive of campaigns' own experiments." *American Political Science Review* 118(4):2021–2039.

Hill, Seth J., James Lo, Lynn Vavreck and John Zaller. 2013. "How Quickly We Forget: The Duration of Persuasion Effects from Mass Communication." *Political Communication* 30(4):521–547.

Hogg, Michael A and Scott A Reid. 2006. "Social identity, self-categorization, and the communication of group norms." *Communication theory* 16(1):7–30.

Holbrook, Allyson L and Jon A Krosnick. 2010. "Social desirability bias in voter turnout reports: Tests using the item count technique." *Public opinion quarterly* 74(1):37–67.

Holman, Mirya R., Monica C. Schneider and Kristin Pondel. 2015. "Gender Targeting in Political Advertisements." *Political Research Quarterly* 68(4):816–829.

Huber, Gregory A and Kevin Arceneaux. 2007. "Identifying the persuasive effects of presidential advertising." *American Journal of Political Science* 51(4):957–977.

Huddy, Leonie, Erin Cassese and Mary Kate Lizotte. 2008. Sources of political unity and disunity among women: Placing the gender gap in perspective. In *Voting the gender gap*. University of Illinois Press pp. 141–169.

Huddy, Leonie, Joshua Billig, John Bracciodieta, Lois Hoeffler, Patrick J Moynihan and Patricia Pugliani. 1997. "The effect of interviewer gender on the survey response." *Political Behavior* 19(3):197–220.

Junn, Jane and Natalie Masuoka. 2024. *Women Voters: Race, Gender, and Dynamism in American Elections*. Cambridge University Press.

Kalla, Joshua L and David E Broockman. 2018. "The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments." *American Political Science Review* 112(1):148–166.

Kam, Cindy D, Allison MN Archer and John G Geer. 2017. "Courting the women's vote: The emotional, cognitive, and persuasive effects of gender-based appeals in campaign advertisements." *Political Behavior* 39(1):51–75.

Kelley, Jr., Stanley and Thad W. Mirer. 1974. "The Simple Act of Voting." *American Political Science Review* 68(2):572–591.

Kelman, Herbert C. 1958. "Compliance, identification, and internalization three processes of attitude change." *Journal of conflict resolution* 2(1):51–60.

Kelman, Herbert C. 2006. "Interests, relationships, identities: Three central issues for individuals and groups in negotiating their social environment." *Annu. Rev. Psychol.* 57(1):1–26.

Klar, Samara. 2013. "The Influence of Competing Identity Primes on Political Preferences." *Journal of Politics* 75(4):1108–1124.

Klar, Samara. 2018. "When common identities decrease trust: An experimental study of partisan women." *American Journal of Political Science* 62(3):610–622.

Klar, Samara and Yanna Krupnikov. 2016. *Independent Politics: How American Disdain for Parties Leads to Political Inaction*. New York: Cambridge University Press.

Krumpal, Ivar. 2013. "Determinants of social desirability bias in sensitive surveys: a literature review." *Quality & quantity* 47(4):2025–2047.

Krupnikov, Yanna, Spencer Piston and Nichole M Bauer. 2016. "Saving face: Identifying voter responses to black candidates and female candidates." *Political Psychology* 37(2):253–273.

Krysan, Maria and Mick Couper. 2003. "Race in the Live and Virtual Interview: Racial Deference, Social Desirability, and Activation Effects in Attitude Surveys." *Social Psychology Quarterly* 66(4):364–383.

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, MA: Harvard University Press.

Levitan, Lindsey C. and Brad Verhulst. 2016. "Conformity in Groups: The Effects of Others' Views on Expressed Attitudes and Attitude Change." *Political Behavior* 38(2):277–315.

Lindsey, Summer. 2023. "Social Preferences: Measuring Private, Public, and Group Preferences through Focus Groups." *Perspectives on Politics* 21(4):1178–1197.

Lodge, Milton, Marco R Steenbergen and Shawn Brau. 1995. "The responsive voter: Campaign information and the dynamics of candidate evaluation." *American political science review* 89(2):309–326.

Lokhorst, Anne Marike, Carol Werner, Henk Staats, Eric van Dijk and Jeff L. Gale. 2013. "Commitment and Behavior Change: A Meta-Analysis and Critical Review of Commitment-Making Strategies in Environmental Research." *Environment and Behavior* 45(1):3–34.

Makse, Todd and Anand E. Sokhey. 2014. "The Displaying of Yard Signs as a Form of Political Participation." *Political Behavior* 36(1):189–213.

McPherson, Miller, Lynn Smith-Lovin and James Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Mutz, Diana C. 1992. "Impersonal Influence: Effects of Representations of Public Opinion on Political Attitudes." *Political Behavior* 14(2):89–122.

Mutz, Diana C. 2006. *Hearing the Other Side: Deliberative versus Participatory Democracy*. New York: Cambridge University Press.

NBC News. 2024. "Trump Campaign Ramps Up Efforts to Win Over Black Women with New Video." https://www.nbcnews.com/politics/2024-election/trump-campaign-black-women-new-video-rcna168516.

Newsweek. 2024. "Fox News Host Says Wife Voting Harris 'Same Thing as Affair'.". Published November 4, 2024.

Noelle-Neumann, Elisabeth. 1974. "The spiral of silence a theory of public opinion." *Journal of communication* 24(2):43–51.

NPR. 2024. "Republicans for Harris? Quiet Women Voters and Secret Ballots in 2024.". Published November 3, 2024.

Pallak, Michael S., David A. Cook and John J. Sullivan. 1980. "Commitment and Energy Conservation." *Applied Social Psychology Annual* 1:235–253.

Paluck, Elizabeth Levy. 2009. "Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda." *Journal of Personality and Social Psychology* 96(3):574–587.

Pickup, Mark, Erik O Kimbrough and Eline A de Rooij. 2022. "Expressive politics as (costly) norm following." *Political behavior* 44(4):1611–1631.

Prentice, Deborah A and Dale T Miller. 1993. "Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm." *Journal of personality and social psychology* 64(2):243.

Slough, Tara. 2024. "Making a Difference: The Consequences of Electoral Experiments." *Political Analysis* 32(4):384–400.

Suhay, Elizabeth. 2015. "Explaining group influence: The role of identity and emotion in political conformity and polarization." *Political Behavior* 37(1):221–251.

Tankard, Margaret and Elizabeth Levy Paluck. 2016. "Norm Perception as a Vehicle for Social Change." *Social Issues and Policy Review* 10(1):181–211.

Tourangeau, Roger and Ting Yan. 2007. "Sensitive questions in surveys." *Psychological bulletin* 133(5):859.

Valentim, Vicente. 2024. "Political stigma and preference falsification: theory and observational evidence." *The Journal of Politics* 86(4):1382–1402.

Washington Post. 2024. "Women Secretly Voting for Harris Sparks Online Buzz.". https://www.washingtonpost.com/style/power/2024/11/04/women-voting-harris-secretly.

White, Ismail K and Chryl N Laird. 2020. *Steadfast democrats: How social forces shape Black political behavior*. Princeton University Press.

Willer, Robb, Ko Kuwabara and Michael W Macy. 2009. "The false enforcement of unpopular norms." *American Journal of Sociology* 115(2):451–490.

Wood, Wendy. 2000. "Attitude change: Persuasion and social influence." *Annual review of psychology* 51(1):539–570.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SVPBAPPEND.pdf