



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN
ECONOMICS

Ciclo 37

Settore Concorsuale: 13/A1 - ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 - ECONOMIA POLITICA

ESSAYS IN GENDER ECONOMICS AND POLITICAL ECONOMY

Presentata da: Federica Esposito

Coordinatore Dottorato

Matteo Barigozzi

Supervisore

Natalia Montinari

Co-supervisore

Nicola Mastrorocco

Esame finale anno 2026

Abstract

This thesis investigates how beliefs, information and narratives shape individual attitudes and policy preferences in domains central to gender equality and social change. Across three essays—two based on experimental evidence and one using historical data—it explores how the provision of information, the correction of misbeliefs, and the exposure to reactionary narratives influence gender attitudes, support for equality-oriented policies, and political alignment.

The first essay examines attitudes toward gender quotas in the labor market. Using a survey experiment among Italian workers and managers ($N = 2,404$), it tests how informational frames emphasizing either demand-side barriers (discrimination and bias) or supply-side constraints (participation, confidence, and role models) affect support for quotas. Results reveal a persistent gap between stated and behavioral support—measured through donations to an NGO promoting gender quotas—and show that supply-side framing increases donation intensity among individuals with less crystallized beliefs about inequality’s origins.

The second essay investigates social norms surrounding coercive and controlling behaviors among young adults in Europe. A preregistered online experiment with 3,900 participants in Italy, Sweden, and Romania reveals that individuals systematically overestimate peer tolerance for such behaviors. Correcting these misperceptions reduces support for coercive actions and reshapes policy preferences regarding gender-based violence, particularly among men and those with larger initial biases.

The third essay turns to the role of political actors, analyzing how reactionary narratives can reshape political cleavages. Leveraging novel historical data on Phyllis Schlafly’s STOP-ERA campaign (1972–1982), it shows that local exposure to anti-feminist mobilization increased Republican vote shares and fostered more conservative gender attitudes, illustrating how moral panics can transform latent cultural divisions into partisan identities.

Essays Included in this Dissertation

1. *“Beliefs about Gender Inequalities, Narratives and Support for Gender Quotas”*
Co-authors: Luca Di Corato, Natalia Montinari
2. *“Young Adults and Gender Based Violence: the Role of Social Norms and Controlling Behaviors”*
Co-author: Elisa Orlandi
3. *“Moral Panics and the Transformation of Political Cleavages: The STOP-ERA Campaign”*
Sole-authored

Beliefs about Gender Inequalities, Narratives and Support for Gender Quotas*

Luca Di Corato[†] Federica Esposito[‡] Natalia Montinari[§]

Abstract

Gender quotas remain a controversial policy tool despite broad agreement on the need to reduce labor-market gender inequality. This paper studies whether information about how quotas work affects support for them, and how such effects depend on pre-existing beliefs about the causes of inequality. We run a *pre-registered* survey experiment among Italian workers and managers ($N = 2,404$), comparing narratives that frame quotas as addressing either demand-side barriers (discrimination and bias) or supply-side constraints (participation, confidence, and role models). While information increases unincentivized, self-reported support for quotas, it does not affect the incentivized decision to donate, revealing a clear *say-do gap*. Conditional on donating, however, the supply-side narrative significantly increases the amount given. Using open-ended responses, we show that information also reshapes how respondents reason about quotas, especially among individuals with diffuse prior beliefs about inequality's causes. We formalize these patterns in a simple framework featuring two complementary mechanisms: misalignment costs that limit responsiveness among individuals with strong priors, and tail-driven effects whereby donors with extreme beliefs disproportionately shape the impact of narratives on costly support, helping explain why the supply-side narrative is most effective on the intensive margin.

*The experiment was pre-registered in the AEA RCT Registry (AEARCTR-0003252). Ethical approval was obtained from the Bioethical Committee of the University of Bologna. We gratefully acknowledge financial support from Ca' Foscari University of Venice (Fondi di Primo Insediamento) and from the PRIN project “*Advancing Gender Equality through Experiments: New Developments and Applications (AGEENDA)*” (CUP P2022R248L.002). We also thank the participants of the Lund University Microeconomics Seminar (September 2024), the Berlin Workshop on Empirical Public Economics: Gender Economics (26-27 September 2024), the VEM - Verona Experimental Meeting (28-29 November 2024), the CEPR Workshop on People's Understanding of and Support for Economic Policies (St. Gallen, Switzerland, 21-22 March 2025), the BEEN meeting, the 2025 Florence Workshop on Behavioural and Experimental Economics (8-10 May 2025) and the Gender Issue and Development Workshop at the University of Reading (3-4 November 2025) for their helpful comments and suggestions. We also thank our research assistants Maria Grazia Pagliarulo, Vlad C. Rotariu, and Elena Barbieri for their excellent work.

[†]Ca'Foscari University of Venice

[‡]University of Bologna

[§]University of Bologna

Keywords: gender quotas, gender inequality, support for policy, survey experiment
JEL codes: D63, D83, J16, J22, J31, J71.

1 Introduction

Despite prolonged efforts over the past decades, gender gaps in labor market outcomes persist in most advanced economies. Women often outperform men in education and enter the labor market better prepared than ever before (Goldin et al., 2006; Kuzmina and Melentyeva, 2021; Bertrand, 2011). Yet, within a few years after graduation and well before motherhood, men are more likely to be employed full-time, earn higher wages, experience steeper career trajectories, and attain top positions (Francesconi and Parey, 2018; Manning and Swaffield, 2008). Gender pay gaps are particularly pronounced at the top of the earnings distribution, and women remain heavily underrepresented in high-status and high-income sectors, occupations, and leadership roles (Blau and Kahn, 2017; Bertrand, 2018). Consistent with this pattern, women hold only about one-third of leadership positions in OECD countries—a figure that has remained strikingly stable over time despite a wide range of policy interventions (OECD, 2023).

This paper studies how support for gender quotas is shaped by informational narratives and pre-existing beliefs, and when expressed approval translates into concrete, costly action.

The persistence of gender gaps at the top of the labor market has motivated a broad set of policy responses aimed at promoting women’s participation and advancement. Guided by international organizations such as the UN, OECD, and the European Union, countries have adopted measures ranging from equal employment opportunity and pay-equality legislation to affirmative action policies, including gender quotas, as well as pay-transparency regulations. Among these initiatives, the OECD (2013) *OECD Gender Recommendation* explicitly calls on member states to address gender inequalities in employment and leadership, including through the use of gender quotas in both the public and private sectors.¹ While implementation varies across countries -ranging from binding quotas (e.g., Norway, France, Spain, and Italy) to softer targets or voluntary guidelines (e.g., the United Kingdom and the United States)- gender quotas have become a central and highly visible policy instrument in the debate over gender equality.

A large body of work documents that board gender quotas are highly effective at changing who holds corporate power. Quotas robustly increase the share of women on boards and are associated with improvements in observable board characteristics, such as higher average education and lower average age of directors, without systematically harming firm performance or shareholder value (Ferrari et al., 2016). Recent meta-analytic evidence pooling quasi-experimental evaluations across multiple countries confirms very high compliance rates and finds no systematic negative effects of quota-induced changes in board composition on firm performance or stock-market outcomes (Acutis et al., 2024).

At the same time, early evidence suggested that spillovers beyond corporate boards (particularly to broader gender gaps within firms and along the career ladder) were

¹More recently, the European Parliament adopted a directive to improve gender balance among directors of listed companies, requiring large EU firms to achieve at least 40% representation of the under-represented sex among non-executive directors, or 33% among all directors, by June 30, 2026 (European Commission, 2022).

limited. More recent contributions, however, paint a more complex picture. Using Italian data, [Zaccaria et al. 2024](#) document sizeable horizontal spillovers through interlocking directorates, whereby board gender quotas in listed firms increase female representation also in connected, non-regulated firms. Complementing this evidence, [Falconieri et al. 2025](#) show that board quota reforms generate spillovers to executive leadership, increasing the appointment of female CEOs in private firms operating in exposed industries, through channels related to visibility and salience rather than direct regulation. Together, these findings suggest that quotas may affect gender inequality beyond corporate boards through indirect and informational mechanisms.

Despite these documented effects, quotas remain a politically contested policy instrument. While mandatory quotas have accelerated women’s representation on corporate boards, sustaining these gains over time remains challenging, and concerns persist about unintended consequences and limited effects on broader career pipelines. For instance, [Rigolini and Huse 2021](#) show that higher female board shares in Italy led to a concentration of directorships among a small group of women—the so-called “golden skirts” phenomenon—while [Seierstad and Huse 2017](#) find mixed evidence on whether firms genuinely expanded female participation or instead adjusted board structures to comply with regulations.

More fundamentally, support for gender quotas often weakens when policies move from abstract principles to concrete instruments. Proponents view quotas as effective tools to counter discrimination and stereotypes, enhancing women’s empowerment, human capital, and firm performance ([Conde-Ruiz and Profeta, 2015](#)). Critics, by contrast, argue that under-representation primarily reflects individual choices rather than discrimination and warn that quotas may undermine meritocratic selection. These competing narratives suggest that the political sustainability of quota policies depends not only on their effectiveness, but also on how they are framed, understood, and evaluated by the public. Consistent with this view, both theoretical and empirical work emphasizes that the success of gender quotas depends on contextual support, perceived fairness, and the narratives surrounding their introduction ([Arve, 2023](#); [Faniko et al., 2017b](#)).

We conduct a pre-registered survey experiment to study how informational narratives about gender quotas shape individual support for such policies, accounting for prior beliefs about the magnitude and causes of gender inequalities in the labor market. The sample consists of 2,404 Italian workers and managers—individuals directly exposed to potential quota effects and likely to hold informed views on affirmative action. The experiment features three information treatments. The *demand* treatment frames quotas as tools to counter discrimination and biases; the *supply* treatment emphasizes constraints such as underconfidence and the lack of role models; and the *info-causes* treatment presents both types of causes without referencing specific policies. A fourth group serves as a control and receives no information.

We study two main outcomes: unincentivized *self-reported support for gender quotas* and an incentivized measure of *behavioral support*, proxied by participants’ willingness to donate part of a potential 500€ lottery prize to an NGO promoting quotas, with a 50% matching contribution by the research team. While all information treatments

increase stated support—most strongly under the demand-side narrative—they do not affect the decision to donate, revealing a clear say-do gap. Conditional on donating, however, exposure to the supply-side narrative significantly increases the amount given.

Beyond shifting levels of support, information also affects how respondents *reason* about quotas. Using coded open-ended responses, we show that narratives reshape the content and structure of justifications, broadening the set of arguments invoked, particularly among respondents with diffuse prior beliefs about the causes of gender inequality. By contrast, respondents with specific priors—such as work-life balance constraints, women’s ability, or discrimination—exhibit weaker or no responses to the narratives. These patterns are consistent with the conceptual framework developed in Section 8, which features two complementary mechanisms. First, donors experience a disutility when the communicated narrative is misaligned with their prior beliefs, implying that information primarily affects respondents with diffuse or weakly formed priors. Second, the agency’s optimal message is driven not by the average donor but by those in the tails of the belief distribution, whose contributions are most sensitive to ideological alignment and misalignment. Together, these mechanisms help explain why the supply-side narrative is uniquely effective in increasing donated amounts in our setting.

Our results are robust to a range of pre-treatment checks, including attrition and self-selection, respondents’ prioritization of gender equality relative to other strategic challenges, and awareness of the existence and magnitude of gender inequalities [Settele, 2022](#). Our study complements recent work linking beliefs to policy demand—most notably [Settele, 2022](#)—while extending it along several dimensions. Whereas [Settele, 2022](#) examine how information about the magnitude of the gender wage gap affects support for affirmative action in a relatively abstract policy setting, we study a concrete and contested instrument: gender quotas for corporate boards. We further move beyond information about *how large* inequalities are to examine how narratives about *why* they persist—demand-side versus supply-side explanations—shape support for this policy. Finally, by combining unincentivized attitudes with an incentivized behavioral outcome—a costly donation to an NGO promoting quotas—we assess not only whether information shifts approval, but also how it affects the *strength* of support and the resulting say-do gap in a context where respondents are directly exposed to potential quota effects.

Our evidence indicates that alignment between informational frames and prior causal beliefs is central to moving from endorsement to action: demand-side narratives primarily boost stated support, whereas supply-side narratives increase donated amounts among respondents with diffuse priors, helping to account for the observed say-do gap.

The remainder of the paper proceeds as follows. Section 2 briefly reviews the relevant literature on support for gender quotas. Section 3 describes the experimental design; Section 4 presents our hypotheses; Section 5 details the data; Section 6 reports the findings; Section 7 discusses potential mechanisms; Section 8 provides a conceptual framework to support the mechanism driving the results; and section 9 concludes.

2 Review of the literature

This paper connects two strands of research. The first examines how beliefs and information shape support for public policies, emphasizing that attitudes toward redistribution and affirmative action depend not only on self-interest but also on fairness perceptions and causal beliefs about inequality (Stantcheva, 2022; Haaland and Roth, 2023; Alesina et al., 2021; Settele, 2022). The second explores the legitimacy and effectiveness of gender quotas, highlighting that their success hinges on public acceptance and perceptions of justice (Arve, 2023; Faniko et al., 2017a; Bush, 2020; Latura and Weeks, 2022; Muriaas and collaborators, 2024; Zehnter and Nater, 2024). While the first literature identifies beliefs as key drivers of policy demand, the second reveals that fairness framing and context strongly condition quota support. Yet, few studies bridge these perspectives by examining how causal beliefs about inequality interact with information about how quotas work, or by linking attitudinal and behavioral measures of support. We next discuss these literatures in greater detail.

Information, beliefs, and policy demand. A large body of work in behavioral and political economics shows that what people believe about *why* inequality exists crucially shapes their preferences for redistribution and corrective policies. Settele (2022) provides direct evidence in the context of affirmative action. In a survey experiment, she examines general affirmative-action policies and shows that individuals who perceive wider gender gaps in opportunities express stronger willingness to endorse such interventions. By combining self-reported attitudes with an incentivized donation measure, she documents a “say-do” gap between verbal approval and costly behavior, demonstrating that beliefs about the *causes* of inequality are proximate drivers of genuine policy demand.

Beyond affirmative action, several studies explore how narratives and information influence support for redistribution. Alesina et al. (2021) show that framing inequality as a product of luck rather than effort increases demand for redistribution, while Stantcheva (2022) models how “causal narratives” structure policy preferences. Haaland and Roth (2023) find that providing information about discrimination or low mobility shifts fairness perceptions and support for equality-enhancing policies. Collectively, this work highlights the causal role of beliefs and information but primarily relies on abstract policy domains and self-reported measures, leaving open how such mechanisms operate when applied to specific, contested interventions such as gender quotas.

Who supports quotas, and why. A second strand of research focuses directly on public attitudes toward gender quotas and seeks to explain who supports them and through which mechanisms. Zehnter and Nater (2024) show that gender differences in quota attitudes arise because women perceive higher levels of discrimination and thus greater policy necessity, whereas men’s opposition is driven by concerns about reverse discrimination and stigma. Muriaas and collaborators (2024) extend this work through a preregistered survey experiment in Norway that varies both the *domain* of quota application—politics, business, or religion—and its *justification*—social-utility versus group-justice arguments. They find

that fairness-based frames are more persuasive than efficiency-based ones and that effects depend on respondents’ prior attitudes, suggesting that framing and prior beliefs interact in shaping support. Evidence from organizational contexts reinforces these findings: Lefley et al. [2024] document widespread skepticism toward corporate-board quotas in the Czech Republic, particularly among men who view them as violating meritocratic norms. At the same time, Bush [2020] show that quotas can enhance external perceptions of legitimacy: in survey experiments in Sweden and the United States, higher women’s representation abroad increased perceived democratic quality and willingness to provide foreign aid. Together, this literature underscores that quota support is multi-dimensional—shaped by fairness beliefs, ideology, gender identity, and contextual framing—but has relied mainly on self-reported attitudes, rarely considering how underlying causal beliefs interact with informational cues or translate into behavioral support.

Gap and contribution. Despite these advances, both literatures leave critical questions open. Research on policy support demonstrates that beliefs and narratives strongly influence preferences, but typically in abstract settings. Work on gender quotas identifies fairness framing and ideology as central correlates of acceptance but seldom examines how prior causal beliefs about inequality condition the response to information, nor whether such beliefs predict real, costly support.

Our study bridges these literatures by focusing on a concrete and contested instrument—gender quotas—and combining three design elements rarely implemented jointly. First, we elicit respondents’ *ex-ante causal beliefs* about why gender inequality persists, capturing both magnitude assessments and open-ended explanations. Second, we randomize *competing informational narratives* that highlight either demand-side (employer discrimination) or supply-side (self-selection) mechanisms. Third, we measure both *stated* and *incentivized* support through attitudinal responses and a real monetary donation task. This design allows us to test whether *narrative -prior alignment*—the consistency between individuals’ initial beliefs and the information they receive—is pivotal for translating abstract endorsement into concrete behavioral action. In doing so, we contribute to understanding how beliefs mediate the effectiveness of information in shaping support for equity-enhancing policies.

3 Experimental Design

This section outlines the experimental design. We first provide an overview of the survey structure and timeline, followed by details on each component.

3.1 Timeline and Overview

Data were collected between March and June 2023 through an online survey (CAWI) administered in collaboration with Scenari Srl², which recruited participants from its

²Scenari Srl (<https://www.scenari.it/>) is an independent Italian company specialized in market research and opinion surveys, with more than thirty years of experience and a consolidated record of

national panel via email invitations³. Before random assignment to one of the four treatment conditions, respondents completed a background questionnaire and a set of belief-elicitation tasks. These included incentivized measures of (i) perceived gender gaps in the labor market, as well as open-ended questions eliciting (ii) beliefs about the causes of gender inequalities and (iii) perceptions of the most effective policies to address them. These pre-treatment measures allow us to account for heterogeneity in pre-existing beliefs about the magnitude and the causes of gender inequalities when estimating treatment effects.

Figure 1 summarizes the overall survey structure, which is discussed in detail in the next subsection.

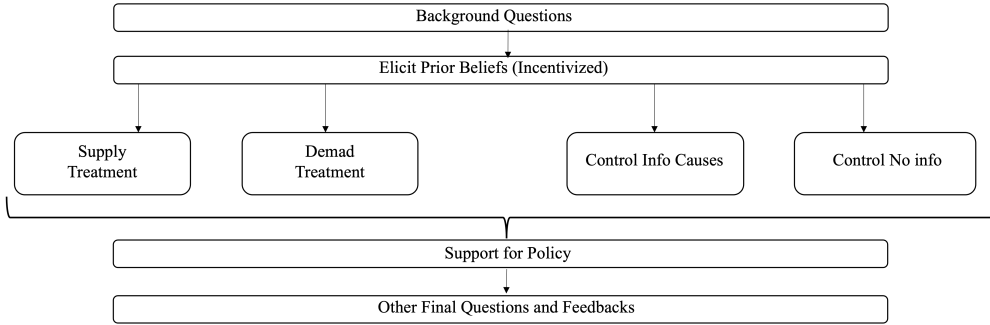


Figure 1: Outline of the survey structure and experimental timeline. The figure illustrates the sequence of stages: background questions, elicitation of prior beliefs, random assignment to one of four treatments, and measurement of post-treatment outcomes.

3.2 The Survey

This subsection describes the structure and content of the survey instrument, detailing the sequence of pre-treatment elicitation, information provision, and post-treatment outcome measurement. Appendix G provides full details on the survey design. Table II summarizes the survey structure and lists the variables elicited at each stage.

collaborations with public institutions and universities, including the Alma Mater Studiorum – University of Bologna, Ca’ Foscari University of Venice, the University of Naples Federico II, the University of Catania, the University of Innsbruck, the University of Turin, and the University of Rome Tor Vergata. The company holds several professional and quality certifications, including ASSIRM, ESOMAR, MSPA, UNI EN ISO 9001:2015, UNI PdR 125:2022, ISO 45001:2018, and EN ISO/IEC 27001:2023. To encourage panel participation, Scenari adopts a point-based incentive system whereby respondents earn points for each completed survey, which can be redeemed for Amazon gift cards.

³To motivate members of the panel to take part in surveys, the company offers incentives by adopting a point-based system. Participants receive points for each survey they complete, depending on the survey length. Every 50 points they can get a 10 Euros Amazon gift card. In the prior beliefs elicitation we provided additional incentives: participants who correctly guessed the answer to the two questions were rewarded with 5 Euros per correct answer with an Amazon gift card paid by the company. Participants were informed that, after the completion of the data collection, 25 participants would be randomly selected and rewarded. Selected participants received the earnings associated with their correct guesses through the company in addition to the participation points awarded to all respondents for completing the survey.

3.2.1 Pre-treatment elicitation

Importance assigned to gender equality. After completing background questions on demographics, employment, and firm characteristics, participants ranked six goals in order of priority for companies over the next five years: (i) promoting innovation, (ii) increasing productivity, (iii) increasing diversity and gender equality, (iv) promoting technological advancement, (v) enhancing teamwork, and (vi) reducing environmental impact. This ranking captures respondents’ baseline prioritization of gender equality before any exposure to the survey content and allows us to test whether this prioritization correlates with survey completion.

Prior belief elicitation. Before treatment assignment, we elicited respondents’ beliefs about three key aspects of gender inequality in managerial positions: its magnitude, its underlying causes, and potential policy remedies. This step is crucial, as participants are likely to hold pre-existing views that may interact with the information treatments.

To assess the perceived magnitude of inequality, respondents answered two incentivized questions: (i) the average wage earned by female managers in Italy for every 100€ earned by male managers aged 30–49 working full-time, and (ii) the share of women CEOs among the 50 largest Italian companies. These measures are precise and difficult to verify online, ensuring that answers reflect genuine beliefs rather than researched values (Settele, 2022). Incentives were awarded to a randomly selected 2% of participants after data collection. Respondents who estimated the gender wage ratio within 2€ or the female CEO share within 2 percentage points of the latest Eurostat and EIGE data received a 5€ bonus per correct answer. This design aims to maintain attention and reduce politically motivated response bias by rewarding accuracy (Bullock et al., 2013; Prior et al., 2015).

To capture prior beliefs about the causes of gender inequality and preferred policy interventions, participants answered two open-ended questions: (i) what they believe causes gender gaps in wages and managerial representation, and (ii) whether they support stronger state intervention to reduce these gaps. Those in favor were asked to specify which policies they consider most effective. These responses allow us to characterize pre-existing beliefs and examine how treatment effects vary with prior views (Stantcheva, 2022).

3.2.2 Information treatments

After completing the pre-treatment elicitation, participants were randomly assigned to one of four groups, three information treatments: *Supply*, *Demand*, and *Info Causes* and one control group: *No Info*.

Participants in the two main treatment groups (*Supply* and *Demand*) watched one of two short, researcher-produced animated videos (Appendix B) explaining how gender quotas can reduce gender inequalities. Each video presented mechanisms through which quotas operate, focusing either on supply-side or demand-side factors.

The *Info Causes* group viewed a video of identical length describing the main causes of gender inequality in managerial positions, but without reference to quotas or policy

interventions. This treatment mitigates potential demand effects arising from exposure to information or video content. The *No Info* group received no information and proceeded directly to the outcome questions, allowing us to isolate any priming effects from watching a video.

The two main treatment videos draw on established research on labor market inequalities. Demand-side explanations emphasize discrimination and bias in hiring or promotion, such as taste-based and statistical discrimination, or implicit bias in evaluating women’s leadership (Beaman et al., 2009; Reuben et al., 2014; Bordalo et al., 2019). Supply-side explanations focus on women’s characteristics and constraints, including heavier family responsibilities, limited leadership aspirations due to few role models, and preferences for less competitive environments (Niederle and Vesterlund, 2007; Günther et al., 2010; Porter and Serra, 2020; Cortés and Pan, 2023).

Both videos share a common introduction (Panels 1–5, Appendix Figure C1) presenting the concept of gender quotas and examples of their adoption across Europe. They then diverge (Panels 6–11, Appendix Figures C2–C3). The *Supply* video links quotas to mechanisms that empower women—such as increasing representation, creating role models, and encouraging investment in human capital—while the *Demand* video highlights how quotas can reduce discrimination by shifting social norms, correcting biased beliefs about competence, and improving fairness in selection processes. The *Info Causes* video covers both demand- and supply-side causes of gender inequality, omitting any discussion of quotas (Appendix Figure C5).

All three videos conclude with a statement noting that the information provided is based on academic research and include a list of references for interested viewers. Participants were required to remain on the video screen for at least 90 seconds (i.e., the full duration of the videos) before proceeding.

3.2.3 Post-treatment elicitation

Self-reported policy demand. After treatment exposure, we measured respondents’ attitudes toward gender quotas in leadership positions, asking whether they consider quotas an effective tool and whether they would support their implementation in their own workplace. We also assessed support for other gender-equality measures, including gender-neutral language, flexible working hours to improve work-life balance, extended paternity leave, and wage transparency. Participants rated the importance of each policy on a four-point scale ranging from “extremely important” to “not important.”

Additional post-treatment questions. To investigate the mechanisms underlying treatment effects, we included several post-treatment measures. First, participants answered an open-ended question explaining the reasons behind their support for or opposition to gender quotas in leadership positions, allowing us to assess how treatments influenced not only levels of support but also the rationale behind respondents’ views.

Next, we introduced a set of mechanism-focused questions. The framing of quotas was randomly varied, presenting them either as increasing the number of women or decreasing

the number of men in organizations. Participants then indicated their agreement with statements suggesting that quotas might lead women (or men) to work less, aspire to higher positions, or reduce teamwork. We also asked respondents to evaluate whether four groups—working men, working women, non-working men, and non-working women—would be advantaged or disadvantaged by quotas.

Finally, to capture potential belief updating induced by the treatments, respondents rated their agreement with statements attributing gender inequalities either (i) to women’s individual effort or (ii) to social norms and historical factors.

Incentivized behavior. In studies on sensitive topics such as gender equality, experimenter demand and social desirability biases are common concerns, as respondents may overstate support for quotas to appear open-minded or to please the experimenter. Although recent evidence suggests such effects may be limited (De Quidt et al., 2018; Mummolo and Peterson, 2019), we included a behavioral measure to mitigate these risks.

Participants were informed that by completing the survey, they were automatically entered into a lottery for a 500€ prize.⁴ Before the draw, respondents indicated (i) whether they would donate any portion of a potential win to an NGO promoting gender quotas, and (ii) if so, the amount (0–500€). Each donated euro was matched by an additional 0.5€ from the research team, creating an incentive to commit to donations immediately rather than hypothetically.

Because even behavioral measures can be affected by self-presentation concerns, we adopted a *forced-response design* (Blair et al., 2015) that allowed participants to decline donation without explicitly saying “no.” Before the donation question, respondents were instructed to consult the third-to-last digit of a banknote or their phone number: if it was 0, 1, or 2, they were required to answer “No”; otherwise, they could freely choose “Yes” or “No.”⁵

Final questions. At the end of the survey, participants answered a short set of concluding questions covering six topics: (i) the perceived importance of gender for their own career prospects; (ii) the presence of workplace initiatives promoting gender equality; (iii) awareness of the Uni Pdr 125:2022 gender-equality certification and whether their company holds it; (iv) perceptions of potential political bias in the study; (v) willingness to donate any bonus earned from the belief-elicitation task if selected for payment; and (vi) interest in receiving additional information about the academic sources cited in the information treatments.

⁴The winner was randomly selected by the survey company after data collection using a standard lottery system.

⁵Exact wording: “In the next question, you will need to answer YES or NO. However, we ask that you follow these instructions: take a banknote or, if you do not have one, think of the third-to-last digit of your phone number. If this digit is 0, 1, or 2, you must answer NO to the next question. In all other cases, you are free to answer YES or NO as you wish.”

4 Hypotheses

This section presents the hypotheses tested in the experiment. All hypotheses were pre-registered prior to data collection on February 15, 2023.⁶ The hypotheses below closely follow the preregistered statements, with minor edits to improve clarity and consistency with the experimental design.

H1. Information provision and support for gender quotas. *Treatments providing information increase support for gender quotas compared to the No Info condition.* This hypothesis builds on prior evidence showing that factual information can affect policy preferences, including in the context of gender inequality and affirmative action (Settele, 2022; Alesina et al., 2021). By providing information about gender inequalities and the functioning of quotas, the treatments are expected to lower informational and psychological barriers, thereby increasing both stated and behavioral support for gender quotas relative to a no-information baseline.

H2. Demand-side versus supply-side framing. *The treatment focusing on demand-side sources of gender inequality increases support for gender quotas more than the treatment focusing on supply-side sources.* This hypothesis reflects the idea that framing gender inequalities as the result of discrimination and structural barriers elicits stronger support for corrective policies than framing them as stemming from supply-side factors, such as women’s preferences or choices (Alesina et al., 2021). Because supply-side explanations are often perceived as reflecting outcomes that are at least partly “chosen” by women, a demand-side narrative is expected to generate higher support for gender quotas.

H3. Role of prior beliefs. *Prior beliefs about the magnitude of gender inequalities affect support for gender-equality policies.* Following Settele (2022), this hypothesis posits that individuals’ pre-existing beliefs shape their support for policy interventions. In our design, prior beliefs are measured along multiple dimensions, including beliefs about the magnitude of gender inequalities, beliefs about their underlying causes, and views on appropriate policy responses. Respondents who perceive larger gender inequalities—and who attribute them to structural or discriminatory factors—are expected to display stronger support for gender quotas than those who perceive inequalities as smaller or primarily driven by individual choices.

H4. Heterogeneity by individual characteristics. *Individual characteristics, such as gender and age, are correlated with treatment effects and with support for gender quotas.* This hypothesis reflects the notion that personal characteristics and lived experiences influence both baseline attitudes toward gender equality policies and responsiveness to information. As a result, treatment effects are expected to vary systematically across demographic groups.

These preregistered hypotheses generate clear empirical predictions about how information provision, framing, prior beliefs, and individual characteristics shape support

⁶Pre-registration DOI: <https://doi.org/10.17605/OSF.IO/FG2XR>.

for gender quotas. We next describe the data and empirical strategy used to test these predictions.

5 Data

This section describes the dataset used in the analysis and assesses whether survey completion is affected by selective attrition related to pre-existing interest in gender-related issues.

5.1 Sample and summary statistics

The final sample consists of 2,404 respondents, all employed individuals aged between 20 and 87. Table 1 reports summary statistics for key demographic characteristics by treatment group, while Table 2 summarizes respondents' job-related characteristics.

As shown in Table 1, women represent slightly more than half of the sample, and respondents are broadly distributed across age groups, regions of residence, and family status. Table 2 indicates substantial heterogeneity in occupational roles, firm types, and job tenure, with representation from both managerial and non-managerial positions and from public, private, and non-profit organizations.

	Demand	Supply	Info Causes	No Info	Total
Women	370	350	334	268	1322
Has children	380	418	354	310	1462
Bachelor's degree or more	320	318	266	256	1160
Age range					
Age 20–25	6	4	13	9	32
Age 26–35	108	63	110	74	354
Age 36–44	193	183	165	140	681
Age 45–54	187	236	189	168	780
Age 55–64	104	125	90	96	415
Age 65–87	31	51	29	31	142
Geographical Area					
Center	115	117	94	66	392
South	183	225	142	203	753
North	325	315	357	244	1241
Observations	629	661	596	518	2404

Table 1: Demographic characteristics of respondents.

5.2 Balance across treatment groups

Appendix Table B1 reports pairwise balance tests across treatment groups for demographic characteristics, while Appendix Table B2 presents corresponding tests for job-related variables. Overall, randomization produced broadly balanced samples. While a small number of pairwise differences emerge for specific characteristics, no systematic patterns

	Demand	Supply	Info Causes	No Info	Total
Job Role					
Manager	75	74	86	85	320
Executive	65	77	27	24	193
Middle Manager	149	161	125	101	536
HR Manager	59	62	75	86	282
Other Managerial Role	28	36	24	19	107
Employees (White Collar)	178	166	158	123	625
Self-Employed	10	12	4	8	34
Employees (Blue Collar / Other)	58	71	88	58	275
Experience					
Average Job Tenure (years)	11.20	12.55	10.89	11.29	11.51
Average Firm Tenure (years)	13.01	14.49	12.21	14.49	13.10
Firm Type					
Public Firm	150	166	156	131	603
Private Firm	432	441	385	322	1580
Non-Profit Firm	47	54	55	65	221
Observations	629	661	596	518	2404

Table 2: Job characteristics of respondents.

are observed. To account for these minor imbalances, all subsequent analyses include standard demographic and job-related controls.

5.3 Attrition and self-selection

We next examine whether survey completion is affected by selective attrition. The overall attrition rate is 12% ($N = 330$ out of 2,734 initial respondents) and is significantly higher among women (16.75%) than among men (5.50%) ($t = -9.05$, $p < 0.001$). Tables [B4](#) in the Appendix report attrition rates by treatment group and present pairwise t -tests. These tests reveal no statistically significant differences in attrition across treatments, including when stratifying the sample by gender. Appendix Table [B5](#) further examines attrition as a function of respondents’ pre-treatment prioritization of gender equality, measured through the ranking task administered before treatment assignment. We find no significant association between the priority assigned to gender equality and survey completion.

The results presented in this sections indicate that survey completion is not driven by self-selection based on pre-existing interest in gender-related issues, supporting the internal validity of the experimental design.

6 Results

We present the results in three steps. First, Section [6.1](#) examines respondents’ perceptions of the magnitude of gender inequalities, their perceived causes, and preferred policy solutions, elicited prior to the information treatments. Second, Section [6.2](#) analyzes the effects of the information treatments on post-treatment outcomes, focusing first on self-

reported support for gender quotas and then on incentivized behavior, with the donation decision and the amount donated considered last. Finally,

6.1 Pre-treatment beliefs about gender inequality

To assess the role of prior beliefs in shaping support for gender quotas we elicited respondents' existing knowledge before exposure to the treatments. Specifically, we elicited participants' beliefs on: (i) the magnitude of gender inequalities in managerial positions (incentivized measure); (ii) the perceived causes of such inequalities; and (iii) potential policy interventions to address them (see Section 6.1.3).

6.1.1 Prior beliefs about the *magnitude* of gender inequalities

We begin by examining respondents' beliefs about the magnitude of gender inequalities in managerial and leadership positions, focusing on two dimensions: the gender wage gap and women's representation at the top of firms.

Gender wage gap. Respondents were asked to estimate how much a woman in a managerial position earns for every 100€ earned by a man in the same role. Based on Eurostat data, the correct value is 55.64€⁷. Participants' mean estimate was 71.48€ (SD = 22.90), corresponding to an inferred gap of 28.52€, which substantially underestimates the true difference. Only 4.62% of participants (111 out of 2,402) provided an estimate within a ± 2 € tolerance window. To quantify the extent of underestimation, we construct an index of misperception, *indexGWG*, defined as the normalized absolute distance between the correct value and each respondent's estimate when the estimate exceeds the true value, and equal to zero otherwise $1 - |y - y^*|/100$, where y is the respondent's answer (on a 0–100 scale) and y^* is the true value⁸. Appendix Table D1 reports pairwise *t*-tests comparing estimated gender wage gaps across treatment groups. No significant differences emerge, indicating that beliefs about the magnitude of the gender wage gap are balanced across treatments. Figure D1 illustrates the distribution of responses by treatment.

Women's representation in top positions. We next consider perceptions of women's representation in top corporate roles. Respondents estimated that, on average, 19.82% (SD = 17.57) of CEOs in Italy are women, whereas the actual share is 2.9%⁹. Only 7.79% of participants (185 out of 2,378) provided an estimate within a ± 2 percentage-point tolerance window.¹⁰ Despite this severe overestimation, we again find no significant differences across treatment groups. Analogously to the wage-gap measure, we define *indexCEO* as the normalized absolute deviation from the true value when the share

⁷Source: *Mean annual earnings by sex, age and occupation – NACE Rev. 2, B–S excluding O* (Eurostat, 2018)

⁸We focus only on deviations that underestimate the true gap. Overestimation occurs in 19.25% of cases ($N = 463/2,404$).

⁹Source: *Largest listed companies: CEOs, executives and non-executives* (EIGE, 2018; Eurostat, 2018)

¹⁰We drop $N = 26$ observations because respondents reported values equal to or above 100%, which are not admissible.

of female CEOs is overestimated, and zero otherwise. Appendix Table D2 reports the corresponding pairwise t -tests, while Figure D2 displays the distribution of responses, confirming the absence of systematic treatment differences.

Overall underestimation of gender inequalities. To capture an overall measure of misperception regarding the magnitude of gender inequalities in leadership positions, we construct the *Underestimation Index*, defined as the average of *indexGWG* and *indexCEO*. The index ranges from 0 (accurate perception, no underestimation) to 1 (maximum underestimation bias)¹¹. Balance tests reported in Appendix Tables D3–D5 confirm that this combined measure is evenly distributed across treatment groups. Figures 3a and 3b report coefficient plots from regressions examining the association between respondents’ characteristics and the Underestimation Index. Women exhibit a significantly higher degree of underestimation, while respondents residing in Southern Italy tend to underestimate less. Turning to job characteristics, individuals employed in large firms (more than 250 employees) display higher levels of underestimation, and employees—relative to managers—also exhibit greater bias.

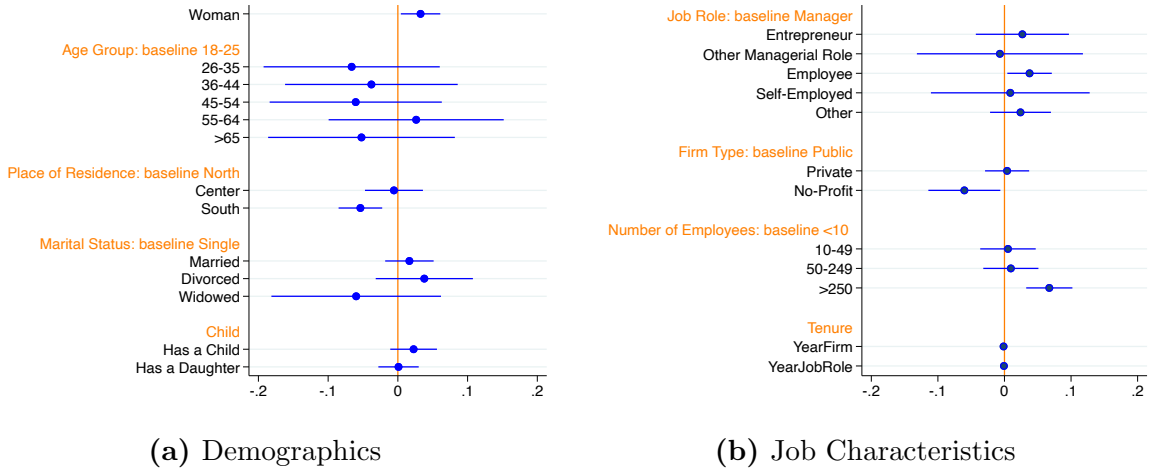


Figure 2: Correlates of the *Underestimation Index*.

Notes: Points show OLS coefficients; bars denote 95% confidence intervals. Each point estimate comes from a *separate* regression (i.e., coefficients are not jointly estimated within a single model). The dependent variable is the *Underestimation Index*, ranging from 0 (accurate perception, no underestimation) to 1 (maximum underestimation of gender inequalities). Robust standard errors are used.

6.1.2 Prior Beliefs about the *causes* of gender inequality

Attitudes toward gender quotas are shaped not only by perceptions of the magnitude of gender inequalities, but also by beliefs about their underlying causes. Individuals who attribute gender gaps to discrimination or structural barriers are generally more supportive of quotas, whereas those who believe disparities stem from women’s preferences, abilities, or “free” choices tend to oppose them. Understanding how individuals conceptualize the

¹¹Underestimation Index equals the average of the CEO and GWG normalized-distance indices and is set to 0 for any respondent whose answer is at or below either true values ($\text{CEO} \leq 2.9$ or $\text{GWG} \leq 55.64$).

sources of gender inequality is therefore central to interpreting both baseline support for quotas and heterogeneous responses to information treatments. To elicit these beliefs, we rely on open-ended questions, which allow respondents to articulate their reasoning in their own words rather than selecting from predefined response options (Ferrario and Stantcheva, 2022). This approach is particularly well suited to capturing potentially multi-dimensional explanations of inequality, which may not fit neatly into standard survey categories.

We analyze respondents’ open-text answers on the causes of gender inequalities in managerial positions and classify them into eight categories derived from the literature on the determinants of gender inequality. The full list of categories is reported in Table 3, with illustrative examples and keywords provided in Appendix Table E2.¹² Respondents appeared to take this task seriously. Answers were typically articulated rather than minimal: the median response length is approximately 9 words (25th percentile: 5 words; 75th percentile: 18 words), and many participants provided multi-clause explanations. This richness allows us to capture underlying reasoning structures rather than relying on surface-level keyword associations.

Text processing and classification of causal beliefs. We follow the text-analysis approach proposed by Ferrario and Stantcheva (2022) to process and classify respondents’ open-ended answers. Responses were first standardized by converting all text to lowercase and removing punctuation, extra spaces, numerical values, and common stopwords (e.g., “and,” “the,” “then”). Words were then stemmed using the Snowball algorithm, which groups different morphological variants under a common root (e.g., “policy” and “policies” → “polic”). This preprocessing yields a cleaned textual representation, denoted by $\hat{d}_i$.

Topic identification was performed using custom dictionaries constructed from the literature on the determinants of gender inequality. For each response, we define a binary indicator equal to 1 if the text contains at least one keyword associated with a given topic, and zero otherwise.¹³ Because a single answer can reference multiple explanations, topic indicators are not mutually exclusive. Consistent with Stantcheva (2022), topic selection was manual rather than algorithmic. This choice reflects the brevity and informal nature of the responses and allows for a more accurate mapping between respondents’ language and the underlying concepts of interest. Manual inspection ensured that keywords captured the intended meaning and reduced the risk of misclassifications due to context-insensitive matches.

Classification of causes. The identified causes are grouped into two broad categories: *generic* and *specific*. The generic category includes broad acknowledgments of gender inequality that do not identify a precise mechanism, such as references to cultural or societal norms or traditional views. Within this group, we define a distinct category, *Cultural*

¹²To ensure data quality, respondents were asked to provide at least five characters.

¹³Responses that could not be classified using the predefined dictionaries were subsequently processed via the ChatGPT API to distinguish between genuinely non-classifiable answers and explicit “no idea” responses.

causes, capturing answers that explicitly attribute inequality to general cultural heritage or prevailing social norms. The specific category comprises responses that point to concrete mechanisms driving gender inequality, including women’s abilities or ambitions, work-life balance constraints, institutional shortcomings, and explicit discrimination or backlash. These subcategories, reported in Table E3, are not mutually exclusive, as respondents frequently mention multiple factors within a single answer. Within the *Women’s ability* category, we distinguish between (i) responses attributing gender gaps to women’s intrinsic traits (e.g., lower ambition or confidence) and (ii) responses referring to how women are perceived by others (e.g., women being viewed as less productive or reliable). Both types of responses are grouped under the same category, as they relate to women’s real or perceived work capabilities and are difficult to disentangle empirically in a consistent manner.

ID	Category	% of responses	Definition
Generic			
1	Cultural causes	28.39%	Acknowledges stereotypes or traditional social norms disadvantaging women, without specifying whether they concern work abilities or social roles.
Specific			
3	Work–Life balance	33.47%	Points to caregiving roles and family responsibilities
2	Women’s ability	28.30%	Mentions women’s (perceived) work capabilities
5	Taste discrimination and reaction	10.37%	Mentions explicit discrimination by men in power or defensive reactions to perceived threats from women.
4	Institutional problems	5.83%	Refers to insufficient institutional or organizational support (e.g., childcare, leave policies).
7	No idea	3.08%	States explicitly not knowing the causes.
6	Denial	1.29%	Denies the existence or salience of gender inequalities.
8	Not classifiable	5.62%	Does not fit any category above.

Table 3: Categories for causes of gender inequality in leadership positions

Notes: Percentages for the four specific categories are computed on the sample with non-missing open-ended responses ($N = 2,402$) and reflect “any mention” (categories are not mutually exclusive). See Appendix Table E1 for examples and keywords.

Table E3 reports the frequency with which each causal category is mentioned, both overall and by treatment group. The most frequently cited explanation for gender inequalities in leadership positions is *work-life balance*, followed by *cultural causes* and *women’s ability*, while institutional problems, explicit discrimination and backlash, and outright denial are mentioned less often. Pairwise *t*-tests by treatment group (Table E3) reveal some statistically significant differences in the distribution of categories across treatments; accordingly, all subsequent analyses control for respondents’ pre-treatment causal beliefs. Correlations across categories, reported in Table E4, indicate that respondents often invoke multiple explanations within the same response, motivating our choice to treat

category indicators as non-mutually exclusive. Clear gender differences emerge (Table E5): women provide more detailed answers overall and are substantially more likely to cite *work-life balance* as a cause (39.6% versus 25.9% among men; $p = 0.000$), whereas men are more likely to offer a generic *cultural cause* explanation (32% versus 25.3%; $p = 0.000$). Regression results in Figure ?? confirm these patterns, showing that being a woman is significantly negatively correlated with invoking a generic explanation, while employment in large firms (more than 250 employees) is positively associated with such responses. Finally, Table 4 links causal beliefs to misperceptions of inequality: respondents who provide generic cultural explanations exhibit significantly higher underestimation, and higher underestimation is also positively correlated with citing *work-life balance* and *taste discrimination and reaction* as causes of the existing inequalities.

Table 4: Regression results: Correlation of Underestimation Index on various causes of gender inequalities

	Cultural Causes	Women’s Ability	Work-Life Balance	Taste discrimination and reaction
Underestimation Index	0.077** (0.026)	-0.006 (0.026)	0.061** (0.028)	0.044** (0.018)
Observations	2,377	2,377	2,377	2,377

Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

At this point, it is informative to characterize respondents’ overall pre-existing views about the causes of gender inequality. Because individual responses can be classified into multiple categories, we first examine how many distinct causes each respondent mentions. Focusing on answers that include at least one of the main specific categories (women’s ability, work-life balance, taste discrimination and reaction, institutional problems), we find that 74.53% of responses mention exactly one category, 21.87% mention two categories (Table E6), only 3.59% mention three or more. This pattern indicates that most respondents articulate relatively focused causal narratives rather than broad or diffuse explanations.

To further isolate interpretable prior views, we restrict attention to respondents who mention exactly one cause and do not invoke the generic cultural category, but instead identify a single specific mechanism ($N = 1,077$). For this subset, beliefs about the causes of gender inequality are cleanly interpretable, as there is no overlap across mechanisms that could blur the distinction between different explanatory frameworks. Among these respondents, 48% attribute gender inequalities primarily to *work-life balance* constraints, a view consistent with a supply-side interpretation emphasizing women’s role as primary caregivers. By contrast, 11.8% attribute inequalities to *taste discrimination and reaction*,

consistent with a demand-side interpretation highlighting employer bias and resistance. This asymmetry mirrors the overall distribution of responses, in which work-life balance emerges as the single most frequently cited cause.

A substantial share of single-cause responses also points to women’s ability (37.1%). However, this category is inherently heterogeneous and difficult to classify unambiguously. While in a few cases respondents clearly refer to external stereotypes or explicit discrimination against women’s capabilities, in many others the wording blurs the distinction between perceived ability, social norms, and individual behavior. Typical responses refer to beliefs about women’s competence, lack of confidence in women by employers or society, women’s self-confidence, or stereotypical traits attributed to women, often combining these elements within the same sentence.¹⁴ As a result, attributing these responses to either supply-side or demand-side mechanisms would require strong assumptions and substantial researcher interpretation. For this reason, we refrain from using the women’s ability category for inference when distinguishing between supply-side and demand-side explanations.

Taken together, these patterns suggest that, prior to treatment exposure, the modal interpretation in our sample is closer to a supply-side explanation of gender inequality. Respondents more often attribute disparities to caregiving responsibilities and traits perceived as incompatible with leadership positions than to explicit discrimination or institutional barriers. This orientation is consistent with the Italian context, where gender norms assigning primary caregiving responsibilities to women remain salient in public discourse, media, and everyday workplace practices [Saraceno, 2020]. Italy also exhibits one of the lowest female labor force participation rates in Western Europe [Eurostat, 2023], large motherhood penalties in employment and working hours [Del Boca and Locatelli, 2022], and comparatively short paternity leave entitlements [OECD, 2023]. These institutional features align with persistent gender asymmetries in unpaid care work and household time allocation. For instance, according to the EIGE Gender Equality Index [for Gender Equality, 2022], Italian women spend approximately five hours per day on unpaid domestic and care work, compared to about two hours for men—a pattern also documented by the ISTAT Time Use Survey [ISTAT, 2019].

6.1.3 Views on Policies

After eliciting respondents’ perceived *causes* of gender inequality, we next assess their prior preferences regarding *policy* responses. Specifically, we examine both whether respondents

¹⁴Illustrative responses include: (i) “There is little trust in women; they are perceived as more incapable and emotional (which is not true)”; (ii) “Prejudices regarding women’s managerial abilities, together with the objective difficulties women face in managing family responsibilities”; (iii) “Gender prejudice, a male-dominated managerial culture, and a lack of confidence both in women and among women themselves”; and (iv) “Women are more cautious and conservative traits in economic and financial decision-making.” See Appendix Table E1 for additional examples and the full set of keywords used in the classification.

believe stronger state intervention is needed and which policy instruments they consider most effective. To limit respondent burden, the open-ended question on effective policies was asked only of individuals who endorsed greater state intervention. Accounting for these pre-treatment policy views is important for isolating treatment effects on policy demand, especially in the case of gender quotas, which some respondents may view as a form of “reverse discrimination,” as discussed in the introduction. As in the previous subsection, we code free-text responses into nine policy categories (see Table 5), with illustrative examples and keywords reported in Appendix Table F1.¹⁵ We apply the same text-processing pipeline and a binary coding rule (equal to one if a category is mentioned and zero otherwise). When gender quotas are explicitly referenced, we further classify the stance as pro- or anti-quota using the ChatGPT API.¹⁶

Appendix Table F2 reports category frequencies by treatment group and overall. The most frequently advocated policies target workplace *discrimination*, followed by measures aimed at *cultural change*. Overall, 10% of respondents (216 out of 2,161) express support for gender quotas, while 1.39% (30 out of 2,161) explicitly oppose them; we find no significant differences across treatments at the pre-treatment stage. Pairwise *t*-tests across treatments, reported in Appendix Table F2, reveal few statistically significant differences. Correlations across policy categories are shown in Appendix Table F3.

Finally, gender differences in policy preferences are reported in Appendix Table F4. Support for *family and motherhood* policies is markedly higher among women (27.5%) than among men (15.8%), with the difference being statistically significant ($p = 0.000$).

ID	Category	Definition
1	Family support policy	Support for motherhood and families.
2	Pro-quota	Favourable reference to gender quotas.
3	Anti-quota	Opposing reference to gender quotas.
4	Anti-discrimination policy	Policies to reduce workplace discrimination.
5	Incentives (non-quota)	Incentives other than quotas (e.g., bonuses, targets).
6	Cultural change	Changing stereotypes via education, media, and social norms.
7	Merit	Emphasis that pay/promotions should depend on merit/skills, not gender.
8	No idea	Not specified / respondent reports not knowing.
9	Not classifiable	Does not fit any category above.

Table 5: Categories for beliefs about policies to reduce gender inequalities

Notes: Categories coded from open-ended responses (binary: 1 if mentioned). Multiple categories may be flagged per response. Pro/anti-quota stance is coded when quotas are mentioned (see Appendix Table F1 for examples/keywords).

¹⁵ Respondents were required to enter at least five characters. Among the 2,161 respondents who supported greater state intervention, 3.47% (75/2,161) answered “no idea” and 15.3% (332/2,161) provided responses that were not classifiable.

¹⁶ Classification procedure as above; pro-/anti-quota tagging applied only when quotas are explicitly mentioned.

To examine how perceived causes of gender inequality relate to the policy respondents propose, we estimate OLS regressions in which the dependent variable is an indicator for mentioning a given type of explanation and the key regressors are dummies for the policy categories mentioned in the same response (Figure 3).

Panel (a) focuses on the generic *cultural causes* category. Respondents who propose anti-discrimination policies and cultural-change interventions are more likely to reference cultural causes, whereas respondents who emphasize family-support policies are markedly less likely to invoke this generic explanation.

Panel (b) focuses on the specific beliefs. The strongest pattern concerns work–life balance: respondents advocating family-support policies are substantially more likely to frame inequality in work–life balance terms, and incentive-based proposals also align positively with this prior belief. By contrast, policy proposals centered on anti-discrimination and merit-based solutions are more consistently associated with explanations that stress misperceptions of women’s ability as a main cause of inequality. Overall, the results indicate that respondents’ policy preferences tend to co-occur with distinct prior beliefs about the causes of gender inequality.

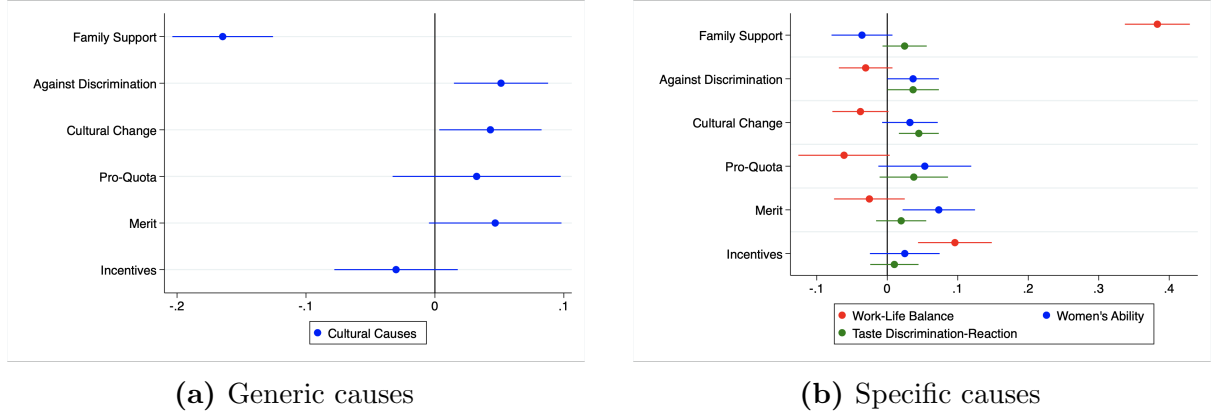


Figure 3: Association between beliefs about causes and policy preferences

Notes: Points show OLS coefficients; bars denote 95% confidence intervals. Each point estimate comes from a *separate* regression (i.e., coefficients are not jointly estimated within a single model). Outcome is an indicator for invoking a given cause; regressors are dummies for policy categories mentioned. Models include demographic and job controls; standard errors are robust.

6.2 Narratives and Support for Gender Quotas: Causal Evidence

As shown above, treatment groups are broadly balanced with respect to pre-treatment beliefs about the magnitude and causes of gender inequality, as well as preferred policy responses. Where minor imbalances arise, we account for them by including appropriate controls in all specifications to preserve internal validity.

We now present the main causal results of the information treatments. First, we examine the effects of the treatments on unincentivized, self-reported support for gender quotas (Section 6.2.1). We then turn to incentivized measures of support, considering in turn (i) respondents’ willingness to donate to an NGO advocating quotas and (ii) the amount donated.

6.2.1 Unincentivized self-reported support for gender quotas

We begin by examining unincentivized self-reported support for gender quotas, measured immediately after the information treatments and prior to the donation task. This outcome may be influenced by social-desirability concerns or experimenter demand; nevertheless, expressed opinions and the narratives underlying them play a central role in shaping attitudes toward gender quotas in both workplace interactions and broader public debate, and constitute the primary outcome in much of the existing literature on quota support (see Section 2). Participants answered two open-ended questions on the perceived effectiveness of quotas and their willingness to support their introduction. We interpret these results alongside the behavioral outcomes. Table 6 reports estimates for the question “*Do you think that quota systems could be a useful tool for reducing inequalities?*” (DV = 1 if “Yes,” 0 otherwise). Across specifications, all information treatments increase self-reported support relative to the No Info condition. The largest effect is observed for the *Demand* treatment (+14.1 percentage points), followed by *Info Causes* (+7.4 pp) and *Supply* (+5.7 pp). Women are also significantly more likely to support gender quotas. Models (1)–(6) progressively introduce additional controls: demographics (2), job characteristics (3), incentivized beliefs about the magnitude of gender gaps and issue prioritization (4), beliefs about causes (generic versus specific) (5), and detailed causes and policy views

derived from open-ended responses (6). Focusing on Model (5), pairwise comparisons indicate that the effect of the *Demand* treatment is significantly larger than that of *Supply* ($p = 0.010$) and *Info Causes* ($p = 0.006$), while the latter two do not differ from each other ($p = 0.969$). These patterns are consistent with the preregistered hypotheses regarding the relative effectiveness of demand- versus supply-side narratives. In the most saturated specification (Model (6)), the indicator for *taste discrimination / reaction* is positive and marginally significant at the 10% level, and indicators for anti-discrimination policies and pro-quota stances are positive and statistically significant. Treatment effects remain stable across specifications, with the exception of *Info Causes*, whose coefficient is no longer statistically significant in the fullest model.

Narratives and ex-post reasoning

To better understand the reasoning underlying respondents’ self-reported support for gender quotas, we analyze the open-ended justifications provided *immediately after treatment exposure*. After indicating whether quotas could address existing inequalities (“Yes/No”), respondents were asked to explain their answer in their own words. Using the coding procedure described in Section 6.1.3 and 6.1.2, we classify these free-text responses into a set of thematic categories, detailed in Appendix Tables G1–G2. In particular, Tables 7 and 8 report estimates of the effects of the information treatments on these *ex post* rationales. For respondents answering “Yes,” we code three main themes: i) *For Women*: quotas are viewed as addressing supply-side constraints (e.g., participation, confidence, role models); ii) *Signaling*: quotas are perceived as helping to reveal women’s abilities that might otherwise be overlooked; iii) *Fairness*: quotas are framed as restoring equality by counteracting long-standing systemic disadvantages. For respondents answering “No,” we code three corresponding themes: i) *Not the Right Instrument*: respondents argue that other policy tools would be more effective than quotas; ii) *Reverse Discrimination*: quotas are viewed as unfairly disadvantaging men; iii) *Unneeded / Backlash*: quotas are considered unnecessary or likely to generate adverse consequences, potentially harming women. Tables 7 and 8 report regression results in

Support for gender quotas						
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (omitted category: No Info)						
Info Causes	0.074*** (0.027)	0.067** (0.030)	0.062** (0.030)	0.061** (0.030)	0.062** (0.030)	0.046 (0.030)
Supply	0.057** (0.027)	0.060** (0.028)	0.064** (0.028)	0.062** (0.028)	0.062** (0.028)	0.063** (0.028)
Demand	0.141*** (0.027)	0.127*** (0.029)	0.129*** (0.029)	0.131*** (0.030)	0.132*** (0.030)	0.107*** (0.029)
Woman		0.047** (0.019)	0.051** (0.021)	0.050** (0.021)	0.049** (0.021)	0.041** (0.021)
Pre-Treatment measures						
Gender Priority				0.001 (0.006)	0.001 (0.006)	-0.002 (0.006)
Underestimation Index				0.000 (0.027)	0.001 (0.027)	-0.021 (0.027)
Pre-treatment beliefs: causes						
Cultural Causes					-0.011 (0.021)	
Women's Ability						0.010 (0.020)
Work-Life Balance						0.002 (0.021)
Taste Discrimination Reaction			-			0.052* (0.029)
No Idea (Causes)						-0.211*** (0.061)
Pre-treatment beliefs: policy responses						
Against Discrimination Policy						0.099*** (0.020)
Pro-Quota						0.267*** (0.021)
Family Support Policy						0.039 (0.025)
Know Quota						0.013 (0.020)
No Idea (Policy)						0.031 (0.062)
Demographics		✓	✓	✓	✓	✓
Job Characteristics			✓	✓	✓	✓
Observations	2,404	2,403	2,403	2,370	2,370	2,370

Notes: Linear probability models. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable equals 1 if the respondent answers “Yes” to “Do you think a quota system could be a useful tool for reducing inequalities?” and 0 otherwise. *Controls:* Demographics — age group, born abroad, macro-region of residence (North, Center, South), college degree, marital status, has children, has a female child. Job characteristics — sector, firm type (public/private/non-profit), firm size (number of employees), job role (manager; entrepreneur; other managerial role; employee—white collar; self-employed; employee—blue collar/other).

Table 6: Effect of information treatments on self-reported support for gender quotas (LPM)

which each specification uses one ex-post justification category—elicited from the open-ended “Yes/No” question on quota effectiveness—as the dependent variable. Regressors include treatment indicators, the *Underestimation Index* from the incentivized belief elicitation tasks, an indicator for whether the respondent views gender equality as a key future challenge for the firm (*Gender Priority*), and dummies capturing prior beliefs about effective policy responses. This empirical setup allows us to assess how information treatments reshape the rationales respondents use to justify their stated positions. Turning first to explanations offered in support of quotas (Table 7), Models (1) and (2) focus on justifications linking quota effectiveness to *supply-side* constraints. Exposure to the *Supply* treatment is associated with a higher likelihood of invoking these arguments, indicating that this narrative shifts perceptions toward viewing quotas as instruments that address supply-side barriers. Models (3) and (4) examine *signaling* rationales—namely, the idea that quotas help reveal women’s abilities that might otherwise be overlooked. Across these specifications, all treatment coefficients are positive and statistically significant. We next consider explanations given for opposing quotas (Table 8). Here, the *Demand* treatment reduces the probability that respondents cite *reverse discrimination* as a justification for opposition. By contrast, the *Info Causes* treatment—providing information about the sources of gender inequality without reference to quotas—raises the likelihood that respondents justify opposition by arguing that quotas are unnecessary or may generate backlash.

Overall, Tables 7 and 8 indicate that information treatments reshape ex-post rationales in systematic ways: supply-side messaging increases the salience of supply-based and signaling justifications in support of quotas, while demand-side messaging reduces the likelihood that opposition is framed in terms of reverse discrimination.

Beyond the *topics* respondents mention, we also examine whether the information treatments affect the *way* respondents reason about gender quotas, as reflected in the structure and articulation of their open-ended justifications. As a first proxy for the structure of reasoning, we consider the number of distinct thematic categories mentioned in each response. Using the coding scheme described above, we count how many different

Ex-post rationales for those who support quotas						
	ForWomen		Signalling		Fairness	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (omitted category: No Info)						
Info Causes	0.001 (0.019)	0.005 (0.019)	0.098*** (0.025)	0.098*** (0.025)	-0.058* (0.031)	-0.057* (0.031)
Supply	0.072*** (0.019)	0.076*** (0.019)	0.115*** (0.024)	0.119*** (0.024)	0.045 (0.029)	0.041 (0.029)
Demand	0.012 (0.019)	0.014 (0.019)	0.189*** (0.027)	0.186*** (0.027)	0.056 (0.032)	0.060* (0.032)
Woman	0.004 (0.014)	0.001 (0.014)	0.042** (0.019)	0.037* (0.019)	0.054** (0.022)	0.058*** (0.022)
Pre-Treatment measures						
Gender Priority	0.004 (0.004)	0.004 (0.004)	0.009 (0.006)	0.008 (0.006)	0.000 (0.006)	0.001 (0.006)
Underestimation Index	0.015 (0.018)	0.011 (0.018)	-0.051* (0.026)	-0.060** (0.026)	0.019 (0.029)	0.033 (0.029)
Pre-treatment beliefs: causes						
Cultural Causes	-0.041*** (0.014)		-0.005 (0.020)		-0.027 (0.023)	-0.027 (0.023)
Women's Ability		0.037** (0.015)		0.039* (0.020)	0.039* (0.023)	0.036 (0.023)
Work-Life Balance		0.062*** (0.016)		0.042** (0.020)	0.030 (0.023)	0.030 (0.023)
Taste Discrimination Reaction		-0.030 (0.020)		0.037 (0.030)	0.059* (0.034)	0.059* (0.034)
No Idea (Causes)		-0.027 (0.028)		-0.144*** (0.033)	-0.204*** (0.044)	-0.147*** (0.054)
Pre-treatment beliefs: policy responses						
Against Discrimination Policy	0.016 (0.014)	0.015 (0.013)	0.081*** (0.018)	0.079*** (0.018)	0.063*** (0.021)	0.067*** (0.021)
Pro-Quota	0.043* (0.025)	0.042* (0.025)	0.097*** (0.034)	0.093*** (0.034)	0.178*** (0.036)	0.184*** (0.036)
Family Support Policy	0.065*** (0.018)	0.050*** (0.018)	0.022 (0.022)	0.005 (0.023)	-0.007 (0.026)	0.003 (0.025)
Know Quota	0.010 (0.014)	0.007 (0.014)	0.019 (0.019)	0.015 (0.019)	-0.018 (0.022)	-0.013 (0.022)
No Idea (Policy)	-0.076*** (0.019)	-0.066*** (0.019)	-0.050 (0.043)	-0.022 (0.044)	-0.108** (0.054)	-0.147*** (0.054)
Demographics	✓	✓	✓	✓	✓	✓
Job Characteristics	✓	✓	✓	✓	✓	✓
Observations	2,370	2,370	2,370	2,370	2,370	2,370

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contains a set of controls. *Demographics:* age group; foreign-born; region (North, Center, South); degree; marital status; has children; has daughter. *Job characteristics:* sector; firm type (public, private, non-profit); firm size; job role. Cultural Causes is mutually exclusive; all other belief categories may co-occur.

Table 7: Explanations for supporting quotas

Ex-post rationales for those who oppose quotas						
	Not Right Instrument (1)	(2)	Reverse Discrimination (3)	(4)	Unneeded / Backlash (5)	(6)
Treatment (omitted category: No Info)						
Info Causes	-0.029 (0.022)	-0.028 (0.022)	0.004 (0.025)	0.004 (0.025)	0.033** (0.014)	0.033** (0.014)
Supply	-0.018 (0.021)	-0.015 (0.021)	-0.028 (0.023)	-0.026 (0.023)	0.002 (0.011)	0.003 (0.011)
Demand	-0.032 (0.022)	-0.032 (0.022)	-0.054** (0.024)	-0.055** (0.024)	0.009 (0.012)	0.009 (0.012)
Woman	-0.026* (0.015)	-0.030** (0.015)	-0.039** (0.018)	-0.042** (0.018)	-0.010 (0.010)	-0.011 (0.010)
Pre-Treatment measures						
Gender Priority	0.003 (0.005)	0.003 (0.005)	0.002 (0.005)	0.002 (0.005)	0.004 (0.003)	0.004 (0.003)
Underestimation Index	0.029 (0.019)	0.029 (0.019)	0.072*** (0.022)	0.071*** (0.022)	0.012 (0.012)	0.013 (0.012)
Pre-treatment beliefs: causes						
Cultural Causes	0.026 (0.016)		0.004 (0.018)		-0.002 (0.009)	
Women's Ability		0.022 (0.016)		0.028 (0.018)		0.016 (0.010)
Work-Life Balance		0.017 (0.017)		0.022 (0.019)		0.009 (0.010)
Taste Discrimination Reaction		-0.021 (0.021)		0.005 (0.026)		-0.016 (0.013)
No Idea (Causes)		-0.074** (0.033)		-0.022 (0.042)		0.014 (0.025)
Pre-treatment beliefs: policy responses						
Against Discrimination Policy	-0.043*** (0.015)	-0.042*** (0.015)	-0.033* (0.017)	-0.034* (0.017)	-0.021** (0.010)	-0.021** (0.010)
Pro-Quota	-0.115*** (0.015)	-0.116*** (0.015)	-0.199*** (0.017)	-0.201*** (0.017)	-0.051*** (0.009)	-0.051*** (0.009)
Family Support Policy	0.021 (0.018)	0.011 (0.019)	-0.032 (0.020)	-0.040* (0.022)	-0.004 (0.011)	-0.006 (0.012)
Know Quota	0.018 (0.015)	0.016 (0.015)	0.049*** (0.017)	0.048*** (0.017)	0.020** (0.009)	0.020** (0.009)
No Idea (Policy)	0.008 (0.046)	0.020 (0.047)	-0.115*** (0.040)	-0.110*** (0.040)	-0.047*** (0.018)	-0.049*** (0.016)
Demographics	✓	✓	✓	✓	✓	✓
Job Characteristics	✓	✓	✓	✓	✓	✓
Observations	2,370	2,370	2,370	2,370	2,370	2,370

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contains a set of controls. *Demographics*: age group; foreign-born; region (North, Center, South); degree; marital status; has children; has daughter. *Job characteristics*: sector; firm type (public, private, non-profit); firm size; job role. Cultural Causes is mutually exclusive; all other belief categories may co-occur.

Table 8: Explanations for opposing quotas

justification categories are invoked within a single message, capturing the extent to which respondents rely on multiple arguments rather than a single rationale. We estimate separate regressions for respondents who report supporting quotas and for those who oppose them. The corresponding results are reported in Appendix Table G4. Among respondents who support quotas, exposure to information treatments—particularly the *Supply* and *Demand* narratives—is associated with a higher number of distinct arguments mentioned, suggesting that these treatments lead supporters to invoke a broader set of considerations when justifying their position. By contrast, among respondents who oppose quotas, we find no systematic evidence that treatments affect the number of categories mentioned. Results are robust to allowing for interactions between treatments and the indicator for citing cultural causes (the generic category), which do not yield qualitatively different patterns G5.

Beyond the *topics* respondents mention, we also examine whether the information treatments affect the *way* respondents reason about gender quotas, as reflected in the structure and articulation of their open-ended justifications. As a first proxy for the structure of reasoning, we consider the number of distinct thematic categories mentioned in each response. Using the coding scheme described above, we count how many different justification categories are invoked within a single message, capturing the extent to which respondents rely on multiple arguments rather than a single rationale. We estimate separate regressions for respondents who report supporting quotas and for those who oppose them. The corresponding results are reported in Appendix Table G4. Among respondents who support quotas, exposure to information treatments—particularly the *Supply* and *Demand* narratives—is associated with a higher number of distinct arguments mentioned, suggesting that these treatments lead supporters to invoke a broader set of considerations when justifying their position. By contrast, among respondents who oppose quotas, we find no systematic evidence that treatments affect the number of categories mentioned. Results are robust to allowing for interactions between treatments and the indicator for citing cultural causes (the generic category), which do not yield qualitatively different patterns (Appendix Table H5).

As a complementary proxy for the articulation of reasoning, we examine the length of respondents’ justifications, measured by the number of words used in the free-form messages provided after the “Quota useful: Yes/No” question. Word counts are computed and regressed on treatment indicators and the same set of pre-treatment controls used in the main specifications. We run separate regressions for those who support quotas and those who do not. The corresponding results are reported in Appendix Table [G6](#). Overall, we find limited evidence that treatments systematically change the length of respondents’ reasoning, with one notable exception: exposure to the *Demand* narrative is associated with longer messages, both among respondents who report supporting or opposing quotas. The results are unchanged when allowing for heterogeneous effects by prior cultural explanations through interactions with the *Cultural Causes* indicator (Appendix Table [G7](#)).

These findings suggest that information treatments primarily affect the *content and breadth* of respondents’ reasoning about gender quotas, while having more limited effects on the overall length of the arguments they provide. Taken together, these analyses speak to how respondents *articulate* and justify their views, rather than to the behavioral mechanisms through which information affects policy support, which we examine next.

6.2.2 Incentivized measures of support

We next examine two incentivized behaviors: the decision to donate to an NGO advocating gender quotas and the amount pledged. Unlike self-reported attitudes, these outcomes are less exposed to experimenter-demand or self-image concerns because they entail a tangible (though contingent) cost: lottery winners must donate the amount they stated. This setting therefore allows us to examine the “say-do” gap—the disconnect between expressed attitudes and actual behavior—and to assess how far stated support for quotas translates into concrete, financially backed action.

Donation decision (extensive margin).

We begin by examining whether exposure to the information treatments affects the likelihood of engaging in any costly support for gender quotas, measured by the decision to donate. Table 9 reports modified-logit estimates for the donation decision, where the dependent variable equals one if the respondent is willing to donate any positive amount to an NGO supporting quotas in the event of a lottery win. The estimator explicitly adjusts for the forced-response design used in the donation item.¹⁷ All specifications include treatment indicators (with No Info as the omitted category) and sequentially add the same sets of controls used in Table 6.

In contrast to the results for unincentivized self-reported support, information treatments have virtually no effect on the willingness to donate. *Gender Priority* is positive and statistically significant, though only marginally so in Model (6). Conditioning on prior beliefs, women are significantly less likely to donate. Prior views also play an important role: indicators for anti-discrimination policy preferences and pro-quota stances are positive and statistically significant, as is *KnowQuota* (pre-treatment knowledge of quotas), highlighting the importance of familiarity and policy preferences in incentivized behavioral support. Results are robust across specifications and in subgroup analyses by gender, job role, and age.

Amount donated (intensive margin)

We next restrict the sample to respondents who choose to donate and examine the causal effect of information on the *amount* donated. Among donors, the average contribution is €87 (SD = 113; range €1–€500), compared to €64 in the control group. Table 10 reports OLS estimates with $\ln(\text{donation})$ as the dependent variable. Each specification includes treatment indicators (with No Info as the omitted category) and sequentially adds the same sets of controls used in the previous tables. Across specifications, the *Supply* treatment exhibits a large and robust effect: conditional on donating, contributions are 66–72% higher than in the control group. This corresponds to an increase of approximately

¹⁷Estimates obtained using the **RRreg** package: Heck et al. 2019.

Decision to Donate (Incentivized)						
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (omitted category: No Info)						
Info Causes	-0.130 (0.175)	0.028 (0.188)	0.027 (0.193)	0.018 (0.194)	0.014 (0.197)	0.030 (0.194)
Supply	0.015 (0.168)	0.017 (0.172)	-0.016 (0.177)	-0.046 (0.178)	-0.045 (0.181)	-0.046 (0.178)
Demand	-0.084 (0.172)	0.082 (0.186)	0.068 (0.193)	0.083 (0.194)	0.043 (0.197)	0.084 (0.194)
Woman		-0.043 (0.123)	-0.207 (0.134)	-0.205 (0.136)	-0.260* (0.140)	-0.207 (0.136)
Pre-Treatment measures						
Gender Priority				0.078*** (0.039)	0.069* (0.040)	0.077** (0.039)
Underestimation Index				-0.235 (0.177)	-0.295 (0.180)	-0.239 (0.177)
Pre-Treatment beliefs: causes						
Cultural Causes					-0.016 (0.202)	
Women's Ability						0.029 (0.141)
Work-Life Balance						0.100 (0.144)
Taste Discrimination—Reaction						0.058 (0.206)
NoIdea (Causes)						-0.125 (0.418)
Pre-Treatment beliefs: policy responses						
Against Discrimination Policy						0.421*** (0.134)
Pro-Quota						0.387* (0.215)
Family Support Policy						0.165 (0.163)
KnowQuota						0.238* (0.138)
NoIdea (Policy)						-0.308 (0.457)
Demographics		✓	✓	✓	✓	✓
Job Characteristics			✓	✓	✓	✓
Observations	2404	2403	2403	2370	2370	2370

Notes: Modified logit estimates accounting for the forced-response design (RRreg). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. *Treatments:* dummies for information treatments; baseline = No Info. *Demographics:* age group; born abroad; macro-region of residence (North, Center, South); college degree; marital status; has children; has a female child (indicator). *Job characteristics:* sector; firm type (public, private, non-profit); firm size (number of employees); job role (manager; entrepreneur; other managerial role; employee—white collar; self-employed; employee—blue collar/other).

Table 9: Decision to donate part of a lottery prize to an NGO supporting gender quotas (modified logit).

€60 relative to the control mean, yielding a predicted average donation of about €124. The *Demand* treatment displays a smaller positive effect (approximately 20–45%), which is weakly significant in the baseline specification and becomes statistically insignificant once controls are included. In the most saturated specification (Model (6)), respondents who attribute gender inequalities to *women’s ability* donate significantly more (significant at the 10% level), and those endorsing *anti-discrimination* policies also exhibit higher donation amounts. Underestimation of the gender gap is negatively associated with donation levels, though the effect is not statistically significant. Finally, the coefficient on *Woman* is negative and statistically significant across all models, implying that women donate roughly 44–52% less than men—consistent with their lower willingness to donate documented above.

Summary and interpretation.

Looking across outcomes, the results diverge sharply. For unincentivized self-reports, both the *Demand* and *Supply* treatments increase stated support for gender quotas (Table 6). By contrast, for incentivized behavior, information treatments do not affect the *decision* to donate (Table 9), and only the *Supply* treatment increases the *amount* donated (Table 10). This divergence highlights the value of combining unincentivized and incentivized measures: self-reports are more susceptible to experimenter demand, self-image, and framing effects, whereas donation outcomes involve tangible (though contingent) costs. This gap is particularly pronounced among women. While women are more likely to express support for quotas in survey responses, they are significantly less likely to donate and, conditional on donating, contribute lower amounts. One plausible interpretation is an income-effect mechanism: donating may be relatively more costly for women, who on average face lower earnings. At the same time, the stronger stated support expressed by women is consistent with prior evidence showing greater support for affirmative action and basic freedoms among women (Bursztyn et al., 2023). The contrast between stated and incentivized outcomes highlights the value of behavioral measures in clarifying how expressed support translates into costly action. To further probe the say–do

	Amount donated (log)					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (omitted category: No Info)						
Info Causes	0.271 (0.271)	0.009 (0.320)	0.018 (0.339)	0.001 (0.339)	0.001 (0.339)	-0.058 (0.329)
Supply	0.660*** (0.248)	0.638** (0.253)	0.683** (0.271)	0.715*** (0.273)	0.715*** (0.274)	0.655** (0.269)
Demand	0.450* (0.251)	0.214 (0.302)	0.218 (0.323)	0.193 (0.325)	0.193 (0.326)	0.185 (0.317)
Woman		-0.445** (0.180)	-0.507** (0.197)	-0.496** (0.200)	-0.497** (0.200)	-0.527** (0.202)
Pre-Treatment measures						
Gender Priority				0.020 (0.054)	0.020 (0.054)	0.018 (0.055)
Underestimation Index				-0.065 (0.247)	-0.065 (0.247)	-0.040 (0.248)
Pre-Treatment beliefs: causes						
Cultural Causes					-0.008 (0.212)	
Women's Ability						0.374* (0.197)
Work-Life Balance						-0.062 (0.214)
Taste Discrimination Reaction						-0.320 (0.351)
No Idea (Causes)						0.532 (0.614)
Pre-Treatment beliefs: policy responses						
Against Discrimination Policy						0.397* (0.211)
Pro-Quota						0.148 (0.312)
Family Support Policy						0.000 (0.221)
Know Quota						0.035 (0.226)
No Idea (Policy)						-1.616** (0.648)
Demographics		✓	✓	✓	✓	✓
Job Characteristics			✓	✓	✓	✓
Observations	416	415	415	412	412	412

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. *Demographics:* age group; born abroad; macro-region of residence (North, Center, South); college degree; marital status; number of children; has daughter (indicator). *Job characteristics:* sector; firm type (public, private, non-profit); firm size (number of employees); job role (manager; entrepreneur; other managerial role; employee—white collar; self-employed; employee—blue collar/other).

Table 10: Effect of treatments on amount donated (log)(range: 1–500 €)

gap, we examine respondents who report supporting quotas post-treatment but decline to donate ex post. This discrepancy may reflect beliefs that donating to the specified NGO is an ineffective way to advance quotas, or it may arise from experimenter-demand and self-image concerns affecting self-reports. In the Online Appendix, we restrict the sample to respondents who stated support for quotas and examine whether information treatments affect the probability of also donating. We find no treatment effects on this extensive margin, confirming that information does not increase the likelihood of donating. Instead, as in the main analysis, treatments operate on the intensive margin by affecting donation amounts. Re-estimating the donation-amount regressions on this restricted sample yields the same qualitative conclusions. Given the reduced sample size, the estimates are less precisely estimated. In the specification without controls, the results remain broadly consistent with the main findings, while in the fully controlled model the *supply* treatment is only marginally significant.

7 Mechanisms

The previous section documented a divergence between unincentivized and incentivized outcomes: information treatments increase stated support for gender quotas but do not affect the decision to donate, while the *Supply* narrative increases donation amounts among those who choose to donate. These patterns motivate an analysis of the mechanisms through which information translates into stated and incentivized support. We focus on belief-based mechanisms and examine whether information treatments operate differently depending on respondents' pre-treatment beliefs about the causes of gender inequality. To this end, we interact treatment indicators with prior-belief categories and estimate these models for all three outcomes: self-reported support (*QuotaUseful*), the donation decision (*Decision to Donate*), and the amount donated (*Donation*). The resulting interaction coefficients are plotted in Figures 4, 5 and 6.

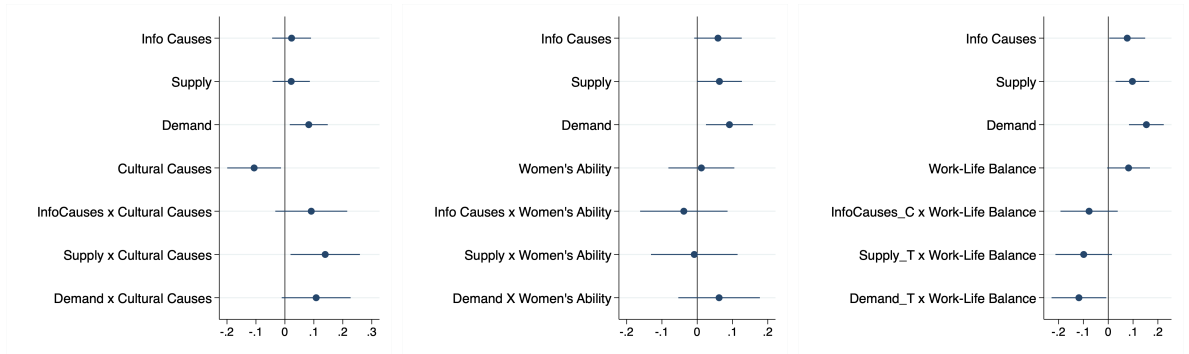


Figure 4: Comparison of belief interactions on self-reported support (*QuotaUseful*)

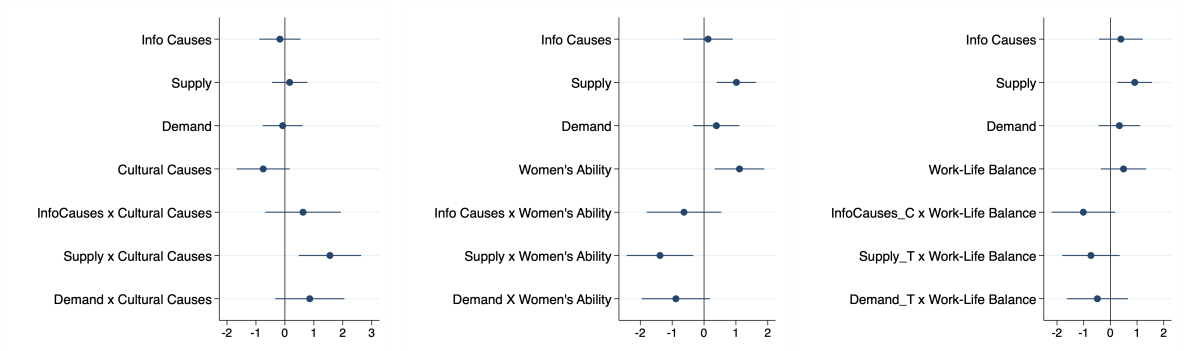


Figure 5: Comparison of belief interactions on $\ln(\text{amount donated})$

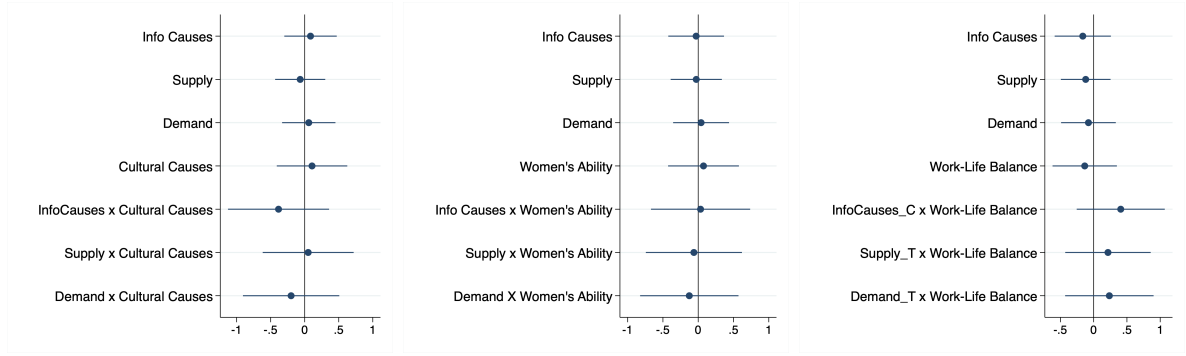


Figure 6: Comparison of belief interactions on decision to donate

Each figure corresponds to one outcome. For each outcome, we estimate three regressions that interact the main prior-belief categories about the causes of inequality with the treatment indicators, controlling for demographics, job characteristics, the Underestimation Index, and prior beliefs about effective policy responses. Consistent with earlier results, we find no significant interactions for the *donation decision*. By contrast, patterns are broadly consistent for *self-reported support* and for the *amount donated*, indicating that prior beliefs systematically moderate treatment effects along these margins. When interacting *cultural causes*—the generic prior-belief category—with treatments, the *Supply* narrative is the only one that increases support among respondents with vague or non-committal beliefs about the causes of inequality. This pattern holds for both outcomes, *QuotaUseful* and *Donation*. In the *No Info* group, respondents with generic beliefs exhibit lower support for quotas (for the donation amount, this effect is marginally significant), indicating that absent information, individuals with diffuse priors are less supportive.

By contrast, in models interacting treatments with more specific prior beliefs—such as *Women's Ability* or *Work-Life Balance*—treatment effects are weak or negative. When respondents hold well-defined priors about the causes of inequality, informational narratives do not increase support and may even backfire.

The results so far help clarifying how information shapes support for gender quotas. Targeted messages are most effective among respondents without strong prior beliefs about the causes of inequality, and the *Supply* narrative emerges as the most effective in increasing both stated support and donation amounts. This pattern suggests that clarifying how

quotas address supply-side constraints can shift policy preferences, particularly among individuals with diffuse or weakly formed priors.

These findings highlight the importance of engaging with pre-existing beliefs about the causes of gender inequality. Targeted information reshapes support for quotas most effectively when respondents' prior beliefs are diffuse, particularly by framing quotas as tools that address women's participation and opportunities. Considering the full set of explanations reported in Section 6.1.2, respondents' prior beliefs about the causes of gender inequality are not evenly distributed. As documented in Section 6.1.2, the most prevalent explanation in the data is a supply-side account, especially work-life balance constraints, while discrimination-based narratives are less frequently invoked. Motivated by the finding that the supply-side narrative is the only one to affect incentivized support, the next section introduces a conceptual framework that explains why the supply-narrative translates into higher donations.

8 The Conceptual Framework

To formalize the intuition emerging from the empirical analysis, we develop a simple conceptual framework that captures how different narratives about the causes of gender inequality influence support for affirmative action. We model the interaction between an agency that communicates on the issue and a pool of potential donors. The agency's goal is to maximize fundraising by strategically choosing which message to convey to potential supporters. We consider two polar narratives: one emphasizing demand-side causes, such as discrimination and bias, and another emphasizing supply-side causes, such as women's preferences and self-selection. Donors hold heterogeneous prior beliefs about the origins of gender inequality. When exposed to a message that conflicts with their priors, they experience a psychological or ideological disutility, even if the information provided could improve their understanding of the issue. This assumption is consistent with the literature on belief utility and cognitive consistency, which highlights that individuals derive utility not only from outcomes but also from holding beliefs that align with their priors (Bénabou

and Tirole, 2011; Yariv, 2002; Bendersky et al., 2022). If this ideological disutility exceeds the informational benefit, donors optimally choose to ignore the message and rely solely on their prior beliefs when deciding how much to donate (see, e.g. Bénabou and Tirole 2016)¹⁸

Let the variable $b \in \{0, 1\}$ denote the message transmitted by the agency, with $b = 0$ representing the supply-side narrative and $b = 1$ the demand-side one. Each donor has a type θ_i drawn from a publicly known continuous distribution function $G(\theta_0)$ over the support $[0, 1]$, with density $g(\theta_0) \geq 0$. The interval $[0, 1]$, captures the heterogeneity in the strength of donors' beliefs about the two competing narratives. This setup parallels a standard Hotelling (1929) spatial competition framework, in which donors' beliefs correspond to preferred locations along an ideological continuum and the distance between a donor's belief and the agency's narrative acts as a cost that shapes willingness to engage.¹⁹

When receiving message b , a donor of type θ_i chooses a donation D to maximize net utility:

$$U(D) = V(D) - C(D; \theta_i, b), \quad i = 1, 2, \quad (1)$$

where $V(D)$ denotes the utility from giving and $C(D; \theta_i, b)$ the associated cost.

We assume that:

$$V(D) = \frac{D^\alpha}{\alpha}, \quad \text{with } \alpha \in (0, 1) \quad (1.1)$$

and

$$C(D; \theta_i, b) = [1 - \lambda \cdot |\theta_i - b| + t_b \cdot (\theta_i - b)^2] \cdot D, \quad \lambda \in (0, 1), \quad t_b > 0. \quad (1.2)$$

In Eq. (1.2), the term $\lambda \cdot |\theta_i - b|$, represents the informational benefit associated with being exposed to a message. This benefit reduces the marginal cost of donating and

¹⁸In the experimental literature, Momen and Ohndorf (2023) and Nam et al. (2013) -among others- shows that individuals sometimes deliberately avoid costless information about payoff and donation consequences when it conflicts with their political predispositions, and that patterns of information avoidance are systematically related to ex-ante ideology.

¹⁹Our model shares a Hotelling-type structure with Mullainathan and Shleifer (2005), where media outlets choose ideological positions to appeal to heterogeneous audiences. Unlike in their framework, where news consumption generates only costs, information in our setting carries intrinsic value but may also entail disutility when misaligned with prior beliefs.

increases linearly (at rate λ) with the distance between prior beliefs and the message received. Conversely, the term $t_b \cdot (\theta_i - b)^2$ captures the ideological disutility that grows quadratically with misalignment between the donor's belief and the message's content.

If the ideological disutility dominates the informational benefit, the donor optimally chooses to ignore the message and bases her donation decision solely on her prior belief. In that case, we set $\theta_i = b$ in Eq. (1.2), which simplifies the cost function to $C(D; \theta_i, \theta_i) = D$.

As a benchmark, consider the case where no communication occurs. In the absence of any message from the agency, donors rely only on their priors ($\theta_i = b$), yielding:

$$D^{NI} = 1, \text{ for any } \theta_i \in [0, 1]. \quad (2)$$

Maximizing $U(D)$ leads to the following result:

Proposition 1. *A donor exposed to message $b \in \{0, 1\}$ donates:*

$$D^*(\theta_i; b) = \begin{cases} [1 - \lambda \cdot |\theta_i - b| + t_b \cdot (\theta_i - b)^2]^{-\frac{1}{1-\alpha}}, & \text{if } |\theta_i - b| < \lambda/t_b, \\ 1, & \text{if } |\theta_i - b| \geq \lambda/t_b. \end{cases} \quad (3)$$

According to Eq. (3), donations peak at $|\bar{\theta} - b| = \frac{\lambda}{2t_b}$, and all donors contribute at least as much as under no communication. When beliefs and the message are perfectly aligned ($\theta_i = b$), $D^*(b; b) = 1$: the message yields no utility but no disutility. If $\lambda > t_b$, the informational benefit dominates for all other types, and $D^*(\theta_i; b) > 1$. Conversely, when $\lambda \leq t_b$, only donors sufficiently close to the message ($|\theta_i - b| \leq \lambda/t_b$) respond to it; other ignore the communication and donate $D^*(\theta_i; b) = 1$.

8.1 The Agency's Problem

The agency chooses the message b that maximizes expected total donations:

$$\max_{b \in \{0, 1\}} F(b) = \int_0^1 D^*(\theta_i; b) \cdot g(\theta_i) \cdot d\theta_i \quad (4)$$

where

$$D^*(\theta_i; 0) = \begin{cases} (1 - \lambda \cdot \theta_i + t_0 \cdot \theta_i^2)^{-\frac{1}{1-\alpha}}, & \theta_i < \lambda/t_0, \\ 1, & \theta_i \geq \lambda/t_0. \end{cases} \quad (4.1)$$

and

$$D^*(\theta_i; 1) = \begin{cases} [1 - \lambda \cdot (1 - \theta_i) + t_1 \cdot (1 - \theta_i)^2]^{-\frac{1}{1-\alpha}}, & \theta_i > 1 - \lambda/t_1, \\ 1, & \theta_i \leq 1 - \lambda/t_1. \end{cases} \quad (4.2)$$

Solving Problem (4) leads to the following result:

Proposition 2. *The agency's optimal message is*

$$b^* = \begin{cases} 0, & \text{if } \Delta \geq 0, \\ 1, & \text{otherwise,} \end{cases} \quad (5)$$

$$\text{where } \Delta = \int_0^1 \left(\frac{dD^*(\theta_i, 0)}{d\theta_i} - \frac{dD^*(\theta_i, 1)}{d\theta_i} \right) \cdot (1 - G(\theta_i)) \cdot d\theta_i.$$

The intuition behind Proposition (2) is straightforward. The relative appeal of the two narratives depends on how individual donations vary across donor types, as reflected in the term $\frac{dD^*(\theta_i, 0)}{d\theta_i} - \frac{dD^*(\theta_i, 1)}{d\theta_i}$ within Δ , weighted by the upper tail of the donor distribution, $(1 - G(\theta_i))$. The weight $(1 - G(\theta_i))$ highlights that donors located in the tails of the distribution play a crucial role in shaping the optimal narrative, as they are more sensitive to ideological alignment and, therefore, to the perceived credibility of the communication. In other words, the agency's choice is not driven by the average donor but by those at the extremes, whose beliefs most strongly amplify or dampen the effect of the message.

When $\lambda \geq t_0, t_1$, this comparison spans the entire interval $[0, 1]$. When $\lambda < t_0, t_1$, only donors who respond positively to the message (those for whom $D^*(\theta_i; b) > 1$) are relevant and, in Eq. (5), the inequality $\Delta \geq 0$ in Eq. (5) can be restated as follows:

$$\int_0^{\frac{\lambda}{t_0}} \frac{dD^*(\theta_i, 0)}{d\theta_i} \cdot (1 - G(\theta_i)) \cdot d\theta_i \geq \int_{1-\frac{\lambda}{t_1}}^1 \frac{dD^*(\theta_i, 1)}{d\theta_i} \cdot (1 - G(\theta_i)) \cdot d\theta_i$$

In this case, the relative size of these “responsive” intervals -where donations exceed the baseline- becomes central in determining the optimal communication strategy.

We allow the parameter t_b to differ across messages to capture variation in the

ideological cost associated with each narrative. Suppose, for instance, that donating under the supply-side message is, *ceteris paribus*, less costly than under the demand-side message, i.e. $t_0 < t_1$. Differentiating $F(1)$ with respect to t_1 yields:

$$\frac{dF(1)}{dt_1} = \begin{cases} \int_0^1 \frac{dD^*(\theta_i, 1)}{dt_1} \cdot g(\theta_i) \cdot d\theta_i < 0, & \text{if } t_0 < t_1 \leq \lambda, \\ \int_{1-\frac{\lambda}{t_1}}^1 \frac{dD^*(\theta_i, 1)}{dt_1} \cdot g(\theta_i) \cdot d\theta_i < 0, & \text{if } \lambda < t_0 < t_1. \end{cases}$$

Thus, as t_1 diverges from t_0 , the likelihood that the demand-side message maximizes expected donations decreases, regardless of the informational value of communication.

Taken together, the model helps explain why the *Supply* narrative is the only one to increase donated amounts in the experiment. The agency’s optimal message is driven not by the average donor but by those in the tails of the belief distribution, whose donations are most sensitive to ideological alignment. Because respondents’ priors are predominantly oriented toward supply-side explanations —most notably work-life balance constraints (Section [6.1.2](#))— a supply-side communication minimizes misalignment for a larger share of donors and therefore maximizes donations, consistent with the experimental evidence.

9 Conclusion

We ran a survey experiment with 2,404 Italian workers to study what shapes support for a highly debated affirmative-action tool: gender quotas. Treatments provided short informational narratives about quota effectiveness framed around either *demand*-side barriers (biases, discrimination) or *supply*-side barriers (participation constraints, confidence, self-selection). Before treatment exposure, we elicited respondents’ priors about the magnitude of gender gaps and, crucially, about their *causes* and the policies best suited to address them, allowing us to account for heterogeneity in pre-existing views.

Across outcomes, we document a clear divergence between unincentivized attitudes and incentivized behavior. Information increases *stated* support for quotas, while the donation decision is largely unchanged; by contrast, among those who donate, the *Supply*

narrative robustly increases the *amount* donated. These patterns motivate our focus on mechanisms that reconcile higher expressed support with limited behavioral movement on the extensive margin but meaningful changes on the intensive margin.

A key contribution of the paper is to show that treatments affect not only *whether* respondents say they support quotas, but also *how* they justify that stance. Using coded open-ended responses, we show that informational narratives change the *structure and content* of reasoning: among quota supporters, exposure to *Supply* and *Demand* increases the number of distinct justification categories invoked, whereas effects are weaker or absent among quota opponents. By contrast, treatments have more limited effects on the overall length of justifications.

The mechanism evidence points to the central role of *prior beliefs about causes*. In particular, treatment effects concentrate among respondents with more *diffuse* priors (the generic “cultural causes” category): for these respondents, the *Supply* narrative is the only one that systematically raises support—and, among donors, donated amounts—while effects are null or even negative when priors are more specific (e.g., women’s ability or work–life balance).

We formalize this intuition in a simple conceptual framework featuring two complementary mechanisms. First, donors experience a disutility when the communicated narrative is misaligned with their prior beliefs, implying that engagement is highest when narratives match priors and can decline when misalignment costs dominate. This mechanism explains why information primarily affects respondents with diffuse or culturally framed beliefs, who lack strong prior commitments and are therefore more receptive to clarifying narratives. Second, the agency’s optimal message is driven not by the average donor but by those in the tails of the belief distribution, whose contributions are most sensitive to ideological alignment and misalignment. This tail-driven mechanism accounts for why the *Supply* narrative performs best on the intensive margin: when beliefs at the extremes are more closely aligned with supply-side explanations—as in our sample—a supply-oriented message maximizes donated amounts, consistent with the experimental evidence.

Overall, our findings show that informational framing interacts with both the speci-

ficity and the orientation of individuals' pre-existing beliefs. Pairing unincentivized and incentivized measures reveals that narratives can broaden and reshape reasoning—and raise stated support—without necessarily moving costly behavior on the extensive margin, while still affecting costly engagement where belief–narrative alignment is strongest. These results underscore the importance of belief-sensitive communication strategies for building durable support for gender-equality policies.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGTP 5 to improve language and readability, with caution. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

References

- C. D. Acutis, A. Weber, and E. Wurm. The effects of board gender quotas: A meta-analysis. *Labour Economics*, 91:102634, 2024. doi: 10.1016/j.labeco.2024.102634. Open access.
- A. Alesina, M. F. Ferroni, and S. Stantcheva. Perceptions of racial gaps, their causes, and ways to reduce them. Technical report, National Bureau of Economic Research, 2021.
- M. Arve. Underrepresentation, quotas and quality: A dynamic argument for reform. *unknown journal / working paper*, 2023. June 2023 version available.
- L. Beaman, R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova. Powerful women: does exposure reduce bias? *The Quarterly journal of economics*, 124(4):1497–1540, 2009.
- R. Bénabou and J. Tirole. Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics*, 126(2):805–855, 2011. doi: 10.1093/qje/qjr002.
- R. Bénabou and J. Tirole. Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–164, 2016.
- C. Bendersky, M. Zeelenberg, and E. Van Dijk. Motivated beliefs and information processing. *American Economic Review: Insights*, 4(2):123–138, 2022. doi: 10.1257/aeri.20200415.
- M. Bertrand. 17-new perspectives on gender. handbook of labor economics, volume 4b, 4 (11), 1543–1590, 2011.
- M. Bertrand. Coase lecture—the glass ceiling. *Economica*, 85(338):205–231, 2018.
- G. Blair, K. Imai, and Y.-Y. Zhou. Design and analysis of the randomized response technique. *Journal of the American Statistical Association*, 110(511):1304–1319, 2015.
- F. D. Blau and L. M. Kahn. The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, 55(3):789–865, 2017.

- P. Bordalo, K. Coffman, N. Gennaioli, and A. Shleifer. Beliefs about gender. *American Economic Review*, 109(3):739–773, 2019.
- J. G. Bullock, A. S. Gerber, S. J. Hill, and G. A. Huber. Partisan bias in factual beliefs about politics. Technical report, National Bureau of Economic Research, 2013.
- L. Bursztyn, A. W. Cappelen, B. Tungodden, A. Voena, and D. H. Yanagizawa-Drott. How are gender norms perceived? Working Paper 31049, National Bureau of Economic Research, 2023. URL <https://www.nber.org/papers/w31049>.
- S. S. Bush. Gender quotas and international reputation. *American Journal of Political Science*, 2020. fill in volume/issue/pages.
- J. Conde-Ruiz and P. Profeta. Quote italiane, un modello che funziona. *La Voce*, 6, 2015.
- P. Cortés and J. Pan. Children and the remaining gender gaps in the labor market. *Journal of Economic Literature*, 61(4):1359–1409, 2023.
- J. De Quidt, J. Haushofer, and C. Roth. Measuring and bounding experimenter demand. *American Economic Review*, 108(11):3266–3302, 2018.
- D. Del Boca and M. Locatelli. Gender gaps, care responsibilities, and labour supply in italy. *XXX Journal*, XX:XXX–XXX, 2022.
- European Commission. Directive (eu) 2022/2381 of the european parliament and of the council on improving the gender balance among directors of listed companies and related measures. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2381>, 2022. Official Journal of the European Union, L315/44, 7 December 2022.
- Eurostat. Mean annual earnings by sex, age and occupation – nace rev. 2, b–s excluding o. https://ec.europa.eu/eurostat/databrowser/view/earn_ses18_04/default/table, 2018. Accessed: October 2025.
- Eurostat. Labour force survey main indicators by sex and age. <https://ec.europa.eu/eurostat/web/lfs/data>, 2023. Accessed: DATE.

- S. Falconieri, M. Ortiz, F. Urzua, and P. Volpin. Board gender quotas and female ceos in private firms. Technical report, 2025.
- K. Faniko, D. Bourguignon, O. Sarrasin, and S. Guimond. “when meritocracy opposes quota policy: How education and policy strength impact opinions”. *Personality and Social Psychology Bulletin*, 43(5):638–651, 2017a.
- K. Faniko, N. Ellemers, B. Derks, and F. Lorenzi-Cioldi. Nothing changes, really: Why women who break through the glass ceiling end up reinforcing it. *Personality and Social Psychology Bulletin*, 43(5):638–651, 2017b.
- G. Ferrari, V. Ferraro, P. Profeta, and C. Pronzato. Gender quotas: Challenging the boards, performance, and the stock market. IZA Discussion Paper 10239, IZA Institute of Labor Economics, Bonn, 2016.
- B. Ferrario and S. Stantcheva. Eliciting people’s first-order concerns: Text analysis of open-ended survey questions. In *AEA Papers and Proceedings*, volume 112, pages 163–169. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2022.
- E. I. for Gender Equality. Gender equality index 2022 – italy. https://eige.europa.eu/modules/custom/eige_gei/app/content/downloads/factsheets/IT_2022_factsheet.pdf, 2022. Accessed: DATE.
- M. Francesconi and M. Parey. Early gender gaps among university graduates. *European economic review*, 109:63–82, 2018.
- C. Goldin, L. F. Katz, and I. Kuziemko. The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic perspectives*, 20(4):133–156, 2006.
- C. Günther, N. A. Ekinici, C. Schwierén, and M. Strobel. Women can’t jump?—an experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, 75(3):395–401, 2010.

- I. Haaland and C. Roth. Beliefs about racial discrimination and support for pro-black policies. *Review of Economics and Statistics*, 105(1):40–53, 2023.
- D. W. Heck, M. Moshagen, and M. D. W. Heck. Package ‘rrreg’. 2019.
- ISTAT. Times of daily life. multiscopo survey on households: Time use. <https://www.istat.it/it/files/2019/05/summary-Times-of-daily-life.pdf>, 2019. Accessed: DATE.
- O. Kuzmina and V. Melentyeva. Gender diversity in corporate boards: Evidence from quota-implied discontinuities. *Available at SSRN 3904886*, 2021.
- E. Latura and W. Weeks. Corporate board quotas and gender equality policies in the workplace. *American Journal of Political Science*, 2022. fill in volume/issue/pages.
- F. Lefley, G. Trnková, and H. Vychová. Views on corporate board gender quotas: Evidence from the czech republic. *Corporate Communications: An International Journal*, 29(2): 123–138, 2024. doi: 10.1108/CCIJ-10-2023-0137.
- A. Manning and J. Swaffield. The gender gap in early-career wage growth. *The Economic Journal*, 118(530):983–1024, 2008.
- K. Momsen and M. Ohndorf. Information avoidance: Self-image concerns, inattention, and ideology. *Journal of Economic Behavior & Organization*, 211:386–400, 2023.
- J. Mummolo and E. Peterson. Demand effects in survey experiments: An empirical assessment. *American Political Science Review*, 113(2):517–529, 2019.
- R. Muriaas and collaborators. Attitudes to gender quotas: Why and where to adjust gender imbalance in leadership. *European Journal of Political Research*, 2024. fill in volume/issue/pages.
- H. H. Nam, J. T. Jost, and J. J. Van Bavel. “not for all the tea in china!” political ideology and the avoidance of dissonance-arousing situations. *PloS one*, 8(4):e59837, 2013.

- M. Niederle and L. Vesterlund. Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101, 2007.
- OECD. Oecd recommendation on gender equality in education, employment and entrepreneurship, 2013. URL <https://www.oecd.org/gender/OECD-Recommendation-Gender-Equality.pdf>.
- OECD. Family database: Pf2.1 parental leave. <https://www.oecd.org/social/family-database.htm>, 2023. Accessed: DATE.
- OECD. Gender equality in a changing world, 2023. URL https://www.oecd.org/en/publications/gender-equality-in-a-changing-world_e808086f-en/.
- C. Porter and D. Serra. Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 12(3):226–254, 2020.
- M. Prior, G. Sood, K. Khanna, et al. You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*, 10(4):489–518, 2015.
- E. Reuben, P. Sapienza, and L. Zingales. How stereotypes impair women’s careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408, 2014.
- A. Rigolini and M. Huse. Women and multiple board memberships: Social capital and institutional pressure. *Journal of Business Ethics*, 169:443–459, 2021.
- C. Saraceno. Gender cultures and care regimes in southern europe. In XXX and XXX, editors, *Handbook on Gender and Social Policy*. Edward Elgar, 2020.
- C. Seierstad and M. Huse. Gender quotas on corporate boards in norway: Ten years later and lessons learned. *Gender Diversity in the Boardroom: Volume 1: The Use of Different Quota Regulations*, pages 11–45, 2017.
- S. Settele. How do beliefs about the gender wage gap affect the demand for public policy? *American Economic Journal: Economic Policy*, 14(2):475–508, 2022.

- S. Stantcheva. Understanding of trade. Technical report, National Bureau of Economic Research, 2022.
- L. Yariv. I'll see it when i believe it: A simple model of cognitive consistency. Cowles foundation discussion paper no. 1352, Cowles Foundation for Research in Economics, Yale University, 2002. URL <https://cowles.yale.edu/sites/default/files/files/pub/d13/d1352.pdf>.
- L. Zaccaria, F. Schivardi, and L. Guiso. Changing the board game: Horizontal spillovers of gender quotas. *Available at SSRN*, 2024.
- M. K. Zehnter and C. Nater. Beyond being beneficiaries: Two mechanisms explain why women have more favourable attitudes towards gender quotas than men. *European Journal of Social Psychology*, 2024. fill in volume/issue/pages.

A Summary statistics, sample balance and attrition

	(1) Demad.T	(2) Supply	(3) Info Causes	(4) No Info	(5) Total	(6) p-value (1) = (2)	(7) p-value (1) = (3)	(8) p-value (1) = (4)	(9) p-value (2) = (3)	(10) p-value (2) = (4)	(11) p-value (3) = (4)
Women	370	350	334	268	1322	0.034	0.325	0.016	0.272	0.679	0.151
Age 20-25	6	4	13	9	32	0.476	0.082	0.245	0.016	0.065	0.596
Age 26-35	108	63	110	74	354	0.000	0.557	0.184	0.000	0.009	0.062
Age 36-44	193	183	165	140	681	0.237	0.249	0.175	1.000	0.802	0.806
Age 45-54	187	236	189	168	780	0.022	0.453	0.325	0.135	0.241	0.797
Age 55-64	104	125	90	96	415	0.265	0.493	0.375	0.073	0.869	0.126
Age 65-87	31	51	29	31	142	0.040	0.956	0.432	0.039	0.247	0.410
Center	115	117	94	66	392	0.786	0.243	0.010	0.361	0.020	0.151
South	183	225	142	203	753	0.056	0.037	0.000	0.000	0.068	0.000
North	325	315	357	244	1241	0.150	0.004	0.124	0.000	0.851	0.000
Has Children	380	418	354	310	1462	0.297	0.717	0.845	0.163	0.235	0.879
Bachelor Degree or More	320	318	266	256	1160	0.321	0.029	0.625	0.217	0.655	0.110
Observations	629	661	596	518	2404						

Table B1: Demographics: Number of observations and balance test, p-value from t-tests

	(1) Demad.T	(2) Supply	(3) Info Causes	(4) No Info	(5) Total	(6) p-value (1) = (2)	(7) p-value (1) = (3)	(8) p-value (1) = (4)	(9) p-value (2) = (3)	(10) p-value (2) = (4)	(11) p-value (3) = (4)
Manager	75	74	86	85	320	0.683	0.195	0.029	0.086	0.009	0.361
Executive	65	77	27	24	193	0.451	0.000	0.000	0.000	0.000	0.935
Middle Mannager	149	161	125	101	536	0.779	0.255	0.087	0.153	0.046	0.542
HR Manager	59	62	75	86	282	1.000	0.073	0.000	0.069	0.000	0.057
Other Managerial Role	28	36	24	19	107	0.411	0.713	0.506	0.239	0.151	0.757
Employees_white collar	178	166	158	123	625	0.196	0.484	0.081	0.572	0.588	0.290
Self Employed	10	12	4	8	34	0.755	0.131	0.951	0.071	0.721	0.159
Employees_blue collar and Other	58	71	88	58	275	0.363	0.003	0.270	0.032	0.804	0.079
Mean of Years in Same Job Role	11.200	12.549	10.892	11.289	11.513	0.018	0.593	0.885	0.005	0.045	0.535
Mean of Years in Same Firm	13.079	14.490	12.212	14.490	13.096	0.027	0.188	0.282	0.000	0.002	0.849
Public Firm	150	166	156	131	603	0.598	0.347	0.572	0.667	0.945	0.737
Private Firm	432	441	385	322	1580	0.452	0.130	0.021	0.430	0.105	0.400
Non-Profit Firm	47	54	55	65	221	0.642	0.267	0.004	0.509	0.013	0.075
Observations	629	661	596	518	2404						

Table B2: Work characteristics: Number of observations and balance test, p-value from t-tests

<i>N</i>	Rank Priority: Gender Equality		t-test	p-value
1286	Supply	Demand		
	3.4909	3.455	0.405	.685
1254	Supply	Info Causes		
	3.491	3.451	-0.451	0.652
1174	Supply	No Info		
	3.491	3.514	-0.246	0.806
1224	Demand	Info Causes		
	3.455	3.451	-0.045	0.964
1144	Demand	No Info		
	3.455	3.513	-0.612	0.541
596	Info Causes	No Info		
	3.451	3.514	-0.655	0.513
	516			

Notes: Sample mean and t-test by treatment for gender equality priority values (1-6). The value indicates in which position the participants ranked gender equality among the other five important goals for their workplace in the next five years. The other five issues were: to promote innovation, increase productivity, promote technological progress, increase teamwork, and promote projects aimed at reducing the environmental impact

Table B3: Rank Priority: Gender Equality by treatment

Attrition			Total	t-test	p-value
<i>N</i>	Supply	Demand			
	90	81	171	0.342	0.736
	751	710	1461		
<i>N</i>	Supply	Info Causes			
	90	81	171	-0.011	0.991
	751	677	1428		
<i>N</i>	Supply	No Info			
	90	78	168	-0.608	0.543
	751	596	1347		
<i>N</i>	Demand	Info Causes			
	81	81	162	0.322	0.747
	710	677	1387		
<i>N</i>	Demand	No Info			
	81	78	159	-0.924	0.356
	710	596	1306		
<i>N</i>	Info Causes	No Info			
	81	78	159	-0.604	0.546
	677	596	1273		

Notes: Attrition represents the number of respondents who abandoned the survey before completing it. The last two columns show the t-test statistic and the p-value for the differences in mean by treatment

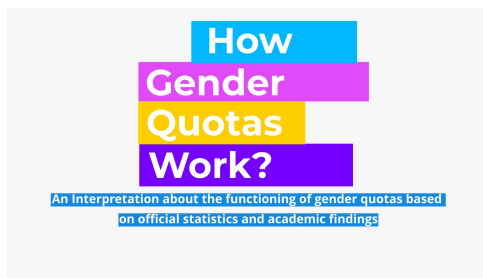
Table B4: Attrition by treatment

	Rank Priority: Gender Equality		Total	t-test	p-value
	Complete	Incomplete			
All Sample					
	3.477	3.514	3.479	-0.276	0.783
<i>N</i>	2398	142	2540		
Demand					
	3.456	3.667	3.468	-0.805	0.421
<i>N</i>	628	39	667		
Supply					
	3.491	3.538	3.494	-0.188	0.851
<i>N</i>	658	39	697		
Info Causes					
	3.451	3.326	3.442	0.524	0.601
<i>N</i>	596	46	642		
No Info					
	3.514	3.611	3.517	-0.608	0.543
<i>N</i>	516	18	534		

Notes: The table contains the sample means for the variable Rank Priority: Gender Equality and the corresponding t-tests and p-values. The value indicates in which position (1-6) the participants ranked gender equality among the other five important goals for their workplace in the next five years. The other five issues were: to promote innovation, increase productivity, promote technological progress, increase teamwork, and promote projects aimed at reducing the environmental impact. The mean at the top represents the entire sample, while the means on the lower rows are computed separately for each treatment and control group.

Table B5: Attrition by Rank Priority: Gender Equality

B Treatment Videos



(a) Panel 1

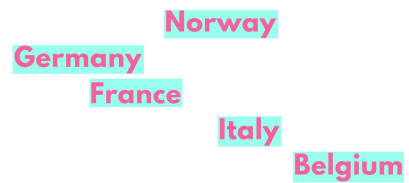
WHAT ARE GENDER QUOTAS?

An instrument for the achievement of gender-balanced participation which establishes a defined number of places for women and/or men

(b) Panel 2



(c) Panel 3



(d) Panel 4



(e) Panel 5

Figure C1: First five panels for both Demand and Supply treatments

Sometimes women tend to have **greater family responsibilities**, so they may find it difficult and costly to dedicate more time to their careers, which can disadvantage them.



(a) Panel S_6

1 Increase the representativeness

Quotas **increase the representation** of women in managerial positions and help **overcome barriers** to advancement



(b) Panel S_7

Inequalities can also be driven by a **lack of women's aspirations** for leadership roles if they do not see other women occupying similar positions.



(c) Panel S_8

2 Supporting ambitions

Female leaders serve as **role models** for other women aspiring to those professional roles



(d) Panel S_9

Finally, according to some studies, women tend to **prefer non-competitive environments** and are less likely to compete for managerial positions



(e) Panel S_10

3 Increase Human Capital

The implementation of policies and role models can **encourage women to invest more in their education, careers, and leadership potential**



(f) Panel S_11

Figure C2: Supply Treatment: Panel 6-11

Women may encounter difficulty in reaching managerial positions because male leaders are often preferred. According to social norms, **leadership is predominantly seen as a masculine trait.**



(a) Panel D_6

1 Reduce Discrimination

Quotas can **overcome** discrimination in the short term and **change** social norms in the long term



(b) Panel D_7

Inequalities can also be determined by a **lack of information** about women's abilities.



(c) Panel D_8

2 Increase the information

Quotas can **correct beliefs** about women in managerial positions and reduce biases.



(d) Panel D_9

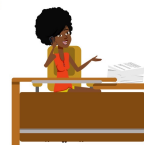
Finally, there are **unconscious mental mechanisms** based on gender, social norms, and cultural values that associate leadership positions with the male figure.



(e) Panel D_10

3 Overcoming implicit biases

Quotas, by reserving positions for women, can **overcome potential unconscious biases** in the selection process.



(f) Panel D_11

Figure C3: Demand Treatment: Panel 6-11



(a) Panel CL1



(b) Panel CL2



(c) Panel CL3

Discrimination

Personal preferences can lead to a **preference for male leaders**. According to social norms, leadership is considered a masculine activity.



(d) Panel CL4

Statistical Discrimination

A different type of discrimination stems from a **lack of information** about the capabilities of female leaders, which can result in selection based on average performance.



(e) Panel CL5

Implicit Bias

Another factor that penalizes women is the possible **automatic mental association** that positions men in leadership roles.



(f) Panel CL6

Figure C4: Control Info Causese: Panel 1-6

The source **2**
 concerns
 women's
 choices

(a) Panel CL7

Prefernces & Costs

Women with **greater family responsibilities** will find it difficult and costly to dedicate more time to their careers.



(b) Panel CL8

Lack of Role Models

Women may not be motivated to aspire to leadership roles unless they see **other women occupying similar positions**.



(c) Panel CL9

Aversion to Competition

According to some studies, women tend to **prefer non-competitive environments** and are less likely to compete for managerial positions



(d) Panel CL10

The information provided comes from academic and scientific studies. If you are interested in obtaining more information, we can provide you with a comprehensive list of the sources used for this video

(e) Panel CL11

Figure C5: Control Info: Panel 7-11

C Incentivized Questions

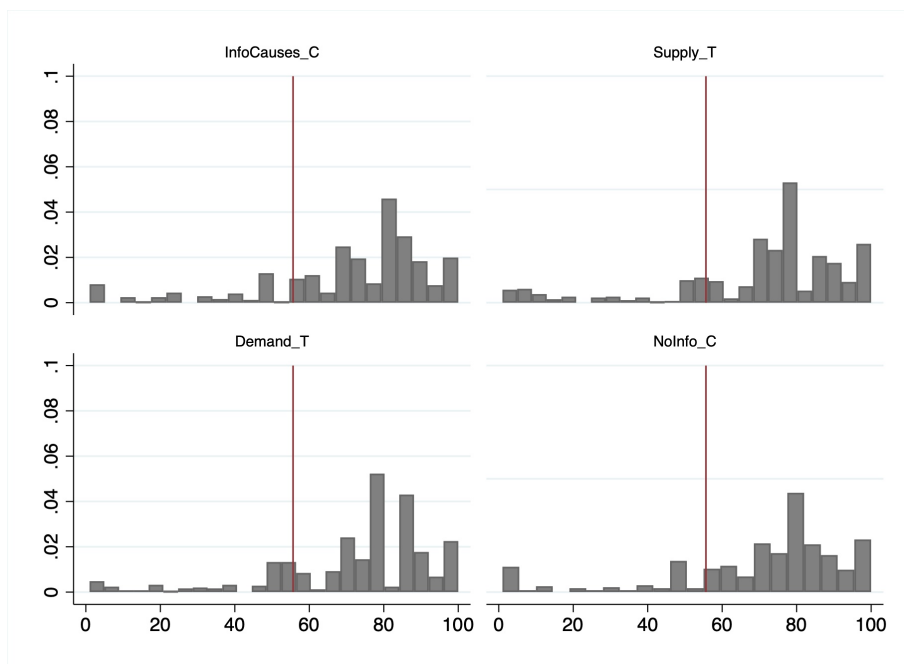


Figure D1: Distributions of prior beliefs about the Gender Wage Gap (as the wage of a female manager for every €100 of a male manager). The red line identifies the right answer: €55.64

Question: *Think about managers (full-time employees aged 30–49) working in Italian firms with more than 10 employees. In your opinion, on average, how much does a female manager with these characteristics earn for every €100 earned by a male manager with the same characteristics?*

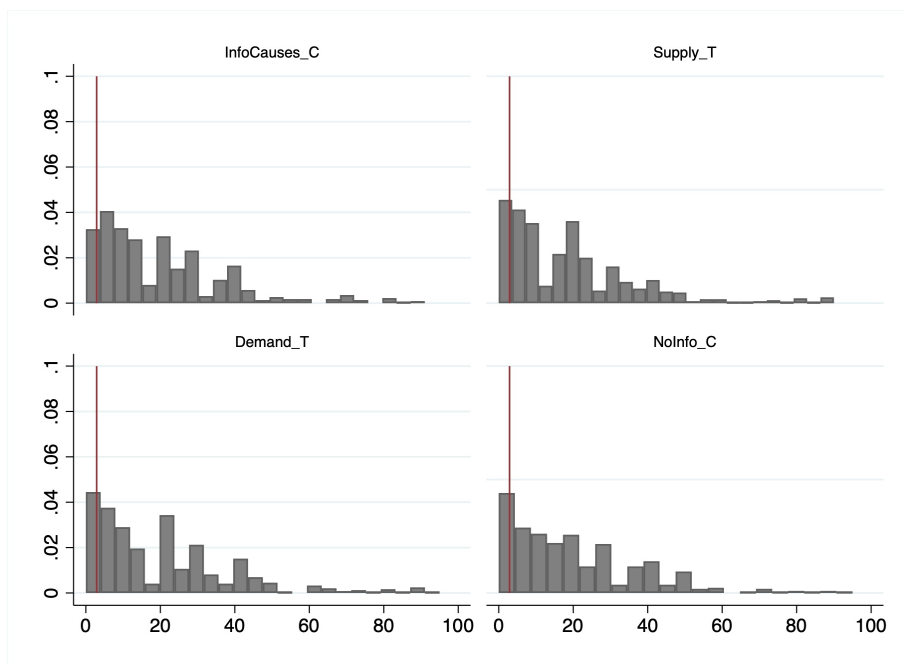


Figure D2: Distributions of prior beliefs about the % of CEO women. The red line identifies the right answer: 2.9%

Question: *In your opinion, what percentage of CEOs of the 50 largest listed companies in Italy are women?*

<i>N</i>	Gender	Wage Gap (GWG)	t-test	p-value
	Supply	Demand		
1289	71.106	72.951	-1.463	0.144
	Supply	Info Causes	t	p
1257	71.106	71.118	0.009	0.993
	Supply	No Info	t	p
1178	71.106	70.596	0.363	0.716
	Demand	Info Causes	t	p
1224	72.951	71.1178	-1.466	0.143
	Demand	No Info	t	p
1145	72.951	70.596	1.765	0.078

Avarage answer to the question: *Think about managers (full-time employees aged 30–49) working in Italian firms with more than 10 employees. In your opinion, on average, how much does a female manager with these characteristics earn for every €100 earned by a male manager with the same characteristics?*

Table D1: Mean and t-test for values of the answers to GWG question (from 1 to 100) by treatment

<i>N</i>	% CEO Women		t-test	p-value
	Supply	Demand		
1272	18.927	20.020	-1.095	0.274
	Supply	Info Causes		
1244	18.927	20.279	1.379	0.168
	Supply	No Info		
1168	18.927	20.185	-1.238	0.216
	Demand	No Info		
1134	20.019	20.185	-0.155	0.877
	Info Causes	Demand		
1210	20.279	20.019	0.253	0.801

Avarage answer to the question: *In your opinion, what percentage of CEOs of the 50 largest listed companies in Italy are women?*

Table D2: Mean and t-test for values of the answers to CEO question (from 1 to 100) by treatment

N	Index GWG		t-test	p-value
1289	Supply	Demand		
	0.610	0.609	0.0065	0.995
1257	Supply	Info Causes		
	0.610	0.602	-0.401	0.688
1178	Supply	No Info		
	0.610	0.600	0.535	0.592
1224	Demand	Info Causes		
	0.609	0.602	-0.3932	0.694
1145	Demand	No Info		
	0.609	0.600	0.528	0.598

IndexGWG is the normalized absolute distance between the true value for the GWG question and the respondent's reported value.

Question: *Think about managers (full-time employees aged 30–49) working in Italian firms with more than 10 employees. In your opinion, on average, how much does a female manager with these characteristics earn for every €100 earned by a male manager with the same characteristics?*

Table D3: Mean and t-test for values of the IndexGWG (from 0 to 1) by treatment

N	Index CEO		t-test	p-value
1272	Supply	Demand		
	0.751	0.737	0.0065	0.995
1244	Supply	Info Causes		
	0.751	0.763	0.741	0.459
1168	Supply	No Info		
	0.751	0.751	0.055	0.956
1210	Demand	Info Causes		
	0.763	0.737	1.615	0.107
1134	Demand	No Info		
	0.737	0.751	-0.787	0.431

IndexCEO is the normalized absolute distance between the true value for the GWG question and the respondent's reported value.

Question: *In your opinion, what percentage of CEOs of the 50 largest listed companies in Italy are women?*

Table D4: Mean and t-test for values of the IndexCEO (from 0 to 1) by treatment

N	Underestimation Index		t-test	p-value
	Supply	Demand		
1271	0.681	0.673	0.621	0.535
	Supply	Info Causes		
1244	0.681	0.683	0.154	0.878
	Supply	No Info		
1167	0.681	0.676	0.389	0.697
	Demand	Info Causes		
1209	0.687	0.673	0.781	0.435
	Demand	No Info		
1132	0.673	0.676	-0.206	0.837

Underestimation Index equals the average of the CEO and GWG normalized-distance indices and is set to 0 for any respondent whose answer is at or below either true values ($\text{CEO} \leq 2.9$ or $\text{GWG} \leq 55.64$).

Table D5: Mean and t-test for values of the Underestimation Index (from 0 to 1) by treatment

	Employment rate	Full-time share	Part-time share
	(1)	(2)	(3)
Underestimation Index	23.315*	-0.004	0.004
	(12.149)	(0.125)	(0.125)
Female education	2.059***	-0.001	0.001
	(0.452)	(0.002)	(0.002)
Centre (ref. North)	-10.380***	0.022	-0.022
	(3.389)	(0.016)	(0.016)
South (ref. North)	-14.306***	0.020	-0.020
	(2.886)	(0.020)	(0.020)
Observations	60	60	60

Notes: Cluster-robust standard errors (region) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All specifications include year fixed effects (omitted from the table). The underestimation index ranges from 0 to 1, where 0 indicates no underestimation and 1 indicates the maximum underestimation, female education is expressed in percentage points.

Table D6: Underestimation index and regional labor market outcomes

	Mean claim 1	Mean claim 2M	Mean claim 3	Mean claim 4	Mean claim 5
	(1)	(2)	(3)	(4)	(5)
Underestimation Index	-0.818	0.051	-1.037**	-1.969**	-0.455
	(0.713)	(0.835)	(0.398)	(0.793)	(0.496)
Female education	0.017	-0.014	0.009	0.028	0.000
	(0.012)	(0.015)	(0.008)	(0.022)	(0.014)
Centre (ref. North)	-0.057	0.025	-0.079	-0.085	-0.057
	(0.078)	(0.139)	(0.065)	(0.202)	(0.102)
South (ref. North)	-0.014	-0.151	0.010	0.351***	0.069
	(0.069)	(0.127)	(0.058)	(0.118)	(0.086)
Observations	60	60	60	60	60

Notes: Cluster-robust standard errors (region) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The underestimation index ranges from 0 to 1, where 0 indicates no underestimation and 1 indicates the maximum underestimation. Female education is expressed in percentage points

Table D7: Underestimation index and regional claim outcomes

D OEQ: Causes of Gender Inequalities

ID	Category	Definition	Examples	Main Keywords
Generic Causes				
1	Cultural causes	References to stereotypes/traditional social norms disadvantaging women, without specifying whether they concern work abilities or social roles.	• “Mentality and culture based on the masculine figure” • “Sexism and the outdated idea that men should dominate” • “Prejudices of society as a whole” • “Entirely cultural problem” • “Effects of a patriarchal society”	Mindset; patriarchy/patriarchal; culture/cultural legacy; male dominance; prejudices; society; biases; stereotypes; sexism; tradition; belief; misogyny; mentality.
Specific Causes				
2	Women’s ability	Mentions women’s (perceived) work capabilities/traits (often undervaluation).	• “Women are considered less at work” • “Belief that women are less perspicacious” • “Household/childcare duties impede performance” (<i>also category 3</i>) • “Women are more cautious and conservative traits in economic and financial decision-making.” • “Prejudices regarding women’s managerial abilities coexist with difficulties related to family responsibilities.” (<i>also category 3</i>) • “Gender prejudice and a male-dominated managerial culture undermine confidence in women and women’s self-confidence.”	Ability; productivity; capacity; ambition; undervalue; underestimate; discrimination; trust; meritocracy.

Table E1: Description of categories for causes of gender inequalities

Table E2: Description of categories for causes of gender inequalities

ID	Category	Definition	Examples	Main keywords
Generic				
1	Cultural causes	References to stereotypes/traditional social norms disadvantaging women, without specifying whether they concern work abilities or social roles.	• “Mentality and culture based on the masculine figure” • “Sexism and the outdated idea that men should dominate” • “Prejudices of society as a whole” • “Entirely cultural problem” • “Effects of a patriarchal society”	Mindset; patriarchy/patriarchal; culture/cultural legacy; male dominance; prejudices; society; biases; stereotypes; sexism; tradition; belief; misogyny; mentality.
Specific				

Table continues on the next page

ID	Category	Definition	Examples	Main keywords
2	Women's ability	Mentions women's (perceived) work capabilities/traits (often undervaluation).	• "Women are considered less at work"	Ability; productivity; capacity; ambition; undervalue; underestimate; discrimination; trust; meritocracy.
			• "Belief that women are less perspicacious"	
			• "Household/childcare duties impede performance" (<i>also category 3</i>)	
			• "Women are more cautious and conservative traits in economic and financial decision-making."	
			• "Prejudices regarding women's managerial abilities coexist with difficulties related to family responsibilities." (<i>also category 3</i>)	
			• "Gender prejudice and a male-dominated managerial culture undermine confidence in women and women's self-confidence."	
3	Work-life balance	Attribution to caregiving roles/family responsibilities constraining careers.	• "Women are seen as homemakers, not career-oriented"	Children; family; pregnancy; leave; caregiving; maternity; domestic duties; absences; time; manage; work-family.
			• "Maternity and domestic chores hinder results" (<i>also category 2</i>)	

Table continues on the next page

ID	Category	Definition	Examples	Main keywords
4	Institutional problems	Insufficient institutional/organizational support (policies, services).	• “Little control from competent bodies” • “Lack of help/support for managing potential maternity” (also 1/3) • “Work does not guarantee adequate assistance” (also 3)	Assistance; policies; support; institutions; laws; state; welfare; unions.
5	Taste discrimination & reaction	Explicit discrimination by men in power and/or defensive reactions to women.	• “Male managers prevent women from entering” • “Fear of losing power by men” • “Fear because women are more skilled”	Leadership positions; disadvantage; obstruct; promote same gender; power; fear.
6	Denial	Denial of the existence/salience of inequalities.	• “In my experience, these inequalities do not exist” • “I don’t think it’s important” • “Maternity is simply an obstacle” (also 3)	“I don’t believe/think”; “there are none”.
7	No idea	Respondent explicitly reports not knowing.	• “I really don’t know” • “I wouldn’t know what to say”	“I don’t know”; “I wouldn’t know”.
8	Not classifiable	Responses that do not fit any category above.	—	—

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Demad	Supply	Info Causes	NoInfo	Total	p(D=S)	p(D=IC)	p(D=C)	p(S=IC)	p(S=C)	p(IC=C)
Cultural Causes	0.275	0.313	0.264	0.280	682	0.134	0.673	0.854	0.057	0.216	0.560
Women's Ability	0.312	0.253	0.305	0.261	679	0.019	0.794	0.068	0.040	0.700	0.121
Work-Life Balance	0.342	0.287	0.345	0.375	804	0.036	0.903	0.250	0.028	0.001	0.308
Institutional Problem	0.062	0.044	0.069	0.060	140	0.146	0.620	0.879	0.053	0.216	0.535
Taste Discrimination Reaction	0.113	0.088	0.120	0.095	249	0.133	0.717	0.315	0.064	0.685	0.182
Denial	0.008	0.014	0.008	0.021	30	0.230	0.928	0.056	0.275	0.432	0.140
No Idea	0.017	0.042	0.032	0.031	74	0.009	0.102	0.137	0.334	0.303	0.334
<i>N</i>	629	661	594	518	2402						

Note that a single answer may be classified into more than one category; therefore, the number of observations (*N*) refers to the total number of answers. The shares are computed as the number of answers containing a given category over the total number of answers within each treatment group. *Cultural Causes* is the only mutually exclusive category.

Table E3: Causes of Inequalities: Share of observations and balance test, p-values from t-tests

	Women's Ability	Work-Life Balance	Institutional Problem	Taste Discrimination and Reaction
Women's Ability	1.000			
Work-Life Balance	-0.038*	1.000		
Institutional Problem	0.061*	0.095*	1.000	
Taste Discrimination and Reaction	0.008	-0.062*	0.050*	1.000

* $p < 0.05$

Table E4: Causes of Inequalities: correlation matrix

	Men	Women	t-test	p-value
Cultural Causes	0.321	0.253	3.680	0.0002
Women's Ability	0.258	0.303	-2.397	0.017
Work-Life Balance	0.260	0.396	-7.066	0.0000
Institutional Problem	0.046	0.068	-2.268	0.0234
Taste Discrimination Reaction	0.093	0.113	-1.609	0.1077
Denial	0.017	0.009	1.666	0.096
No Idea	0.041	0.023	2.549	0.0109
<i>N</i>	2402			

Table E5: Categories Causes: means, t-tests and p-value by gender

<i>Panel A: Responses mentioning exactly one category (N = 1077)</i>	
Category	Share of responses
Women's Ability	0.371
Work-Life Balance	0.480
Taste Discrimination Reaction	0.118
Institutional Problem	0.031
<i>Panel B: Responses mentioning more than one category (N = 368)</i>	
Category	Share of responses
Women's Ability	0.758
Work-Life Balance	0.780
Taste Discrimination Reaction	0.332
Institutional Problem	0.291

Notes: Share of responses in which the category is present. Categories are not mutually exclusive, so shares do not sum to one.

Table E6: Share of responses mentioning each category

	Cultural causes (1)	Work-life balance (2)	Women's ability (3)	Taste discrim. react. (4)
<i>Demographics and background characteristics</i>				
Female	-0.086*** (0.021)	0.107*** (0.021)	0.046** (0.020)	0.014 (0.014)
Has Children	-0.050* (0.027)	0.036 (0.028)	0.030 (0.027)	-0.012 (0.019)
Has a Daughter	0.028 (0.022)	-0.005 (0.023)	-0.041* (0.022)	-0.004 (0.015)
Born abroad	0.051 (0.058)	-0.139*** (0.051)	0.033 (0.057)	0.042 (0.044)
Place of residence (ref. North)				
South	-0.023 (0.022)	-0.037 (0.023)	-0.004 (0.022)	0.003 (0.015)
Center	0.009 (0.028)	-0.002 (0.029)	-0.074*** (0.026)	-0.017 (0.018)
Bachelor's degree or more	0.040* (0.020)	0.019 (0.021)	-0.030 (0.020)	-0.028** (0.013)
Age group (ref. ≤ 25)				
Age 26-35	0.082 (0.086)	-0.089 (0.089)	-0.112 (0.090)	0.061* (0.035)
Age 36-44	0.076 (0.085)	-0.135 (0.088)	-0.148* (0.089)	0.054 (0.035)
Age 45-54	0.073 (0.085)	-0.137 (0.088)	-0.167* (0.089)	0.078** (0.036)
Age 55-64	0.078 (0.087)	-0.200** (0.090)	-0.117 (0.091)	0.099*** (0.038)
Age ≥ 65	0.088 (0.094)	-0.209** (0.095)	-0.146 (0.096)	0.027 (0.041)
Marital Status (ref. Single)				
Married	-0.041 (0.027)	0.066** (0.027)	0.020 (0.027)	0.023 (0.018)
Divorced	-0.042 (0.051)	0.038 (0.051)	0.022 (0.050)	0.028 (0.035)
Widowed	-0.045 (0.082)	0.033 (0.077)	-0.029 (0.085)	-0.019 (0.047)
Observations	2369	2369	2369	2369

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All specifications additionally control for work characteristics (job role, firm size, sector, firm type, and affirmative action at work) and belief measures (underestimation index and gender priority). These variables are omitted from the table for readability (see Table E8).

Table E7: Demographics and background characteristics correlates across beliefs about the causes of gender inequalities

	Cultural causes (1)	Work-life balance (2)	Women's ability (3)	Taste discrim. react. (4)
<i>Work characteristics</i>				
Job Role (ref. Manager)				
Entrepreneur	0.023 (0.048)	-0.043 (0.046)	0.051 (0.050)	0.015 (0.034)
Other Managerial Role	0.023 (0.080)	0.144 (0.090)	0.146* (0.088)	0.028 (0.063)
Employee	-0.015 (0.024)	0.056** (0.026)	0.035 (0.024)	0.006 (0.017)
Self-Employed	-0.081 (0.073)	0.198** (0.086)	0.038 (0.084)	-0.031 (0.041)
Other	0.055* (0.032)	0.034 (0.033)	0.037 (0.032)	-0.035* (0.020)
Number of employees (ref. <10)				
10-49	0.000 (0.029)	-0.048 (0.031)	-0.064** (0.030)	0.018 (0.021)
50-249	0.023 (0.031)	-0.017 (0.032)	-0.025 (0.032)	-0.011 (0.021)
≥250	0.060** (0.028)	0.033 (0.030)	-0.069** (0.028)	-0.015 (0.019)
Firm type (ref. Public)				
Private	0.044* (0.027)	0.025 (0.027)	0.018 (0.026)	-0.008 (0.019)
No profit	0.013 (0.042)	0.005 (0.041)	0.023 (0.043)	-0.008 (0.028)
Affirmative Action at work	0.000 (0.021)	-0.013 (0.022)	0.004 (0.021)	0.003 (0.015)
<i>Prior Beliefs</i>				
Underestimation index	0.079*** (0.026)	0.038 (0.027)	-0.013 (0.027)	0.042** (0.017)
Gender Priority	-0.005 (0.006)	-0.001 (0.006)	0.004 (0.006)	0.010** (0.004)
Observations	2369	2369	2369	2369

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All specifications also control for demographics and background characteristics (included but omitted from the table for readability; see Table E7).

Sector indicators are included in all specifications but omitted from the table for readability.

Table E8: Work characteristics and beliefs across outcomes (full models)

	Employment rate	Full-time share	Part-time share
Cultural causes	4.704 (4.703)	0.085*** (0.026)	-0.085*** (0.026)
Female education	2.315*** (0.482)	-0.001 (0.001)	0.001 (0.001)
Centre (ref. North)	-11.343*** (3.663)	0.031** (0.015)	-0.031** (0.015)
South (ref. North)	-13.623*** (3.408)	0.028 (0.019)	-0.028 (0.019)
Work-life balance	-11.672 (9.803)	-0.113*** (0.038)	0.113*** (0.038)
Female education	2.460*** (0.557)	0.000 (0.001)	-0.000 (0.001)
Centre (ref. North)	-11.687*** (3.516)	0.024 (0.014)	-0.024 (0.014)
South (ref. North)	-13.540*** (3.481)	0.025 (0.020)	-0.025 (0.020)
Women's ability	4.172 (11.957)	-0.130* (0.069)	0.130* (0.069)
Female education	2.343*** (0.508)	-0.002 (0.001)	0.002 (0.001)
Centre (ref. North)	-11.705*** (3.585)	0.019 (0.012)	-0.019 (0.012)
South (ref. North)	-14.338*** (3.243)	0.029 (0.020)	-0.029 (0.020)
Taste discrimination reaction	-5.341 (31.688)	-0.225 (0.180)	0.225 (0.180)
Female education	2.300*** (0.485)	-0.002 (0.002)	0.002 (0.002)
Centre (ref. North)	-11.827*** (3.464)	0.022 (0.014)	-0.022 (0.014)
South (ref. North)	-14.106*** (3.268)	0.018 (0.019)	-0.019 (0.019)
Observations	60	60	60

Notes: Cluster-robust standard errors (region) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All specifications include year fixed effects (omitted from the table). Female education is expressed in percentage points.

Table E9: Regional employment outcomes and beliefs about causes of gender inequalities

	Mean claim 1	Mean claim 2M	Mean claim 3	Mean claim 4	Mean claim 5
Cultural causes	-0.460*** (0.132)	-0.607* (0.308)	-0.382*** (0.088)	-1.241*** (0.375)	-0.409* (0.195)
Female education	0.008 (0.005)	-0.013 (0.011)	-0.002 (0.005)	0.007 (0.013)	-0.004 (0.010)
Centre (ref. North)	-0.053 (0.059)	-0.039 (0.115)	-0.054 (0.054)	-0.089 (0.162)	-0.070 (0.081)
South and Islands (ref. North)	-0.065 (0.060)	-0.207* (0.119)	-0.036 (0.061)	0.215** (0.078)	0.026 (0.081)
Work-life balance	0.739** (0.312)	1.022* (0.537)	0.437 (0.330)	1.727*** (0.513)	0.672*** (0.214)
Female education	-0.001 (0.005)	-0.026** (0.010)	-0.007 (0.005)	-0.015 (0.016)	-0.013 (0.010)
Centre (ref. North)	-0.015 (0.062)	0.011 (0.113)	-0.020 (0.061)	0.017 (0.170)	-0.036 (0.077)
South and Islands (ref. North)	-0.055 (0.059)	-0.196 (0.124)	-0.020 (0.064)	0.253** (0.099)	0.034 (0.076)
Women's ability	0.361 (0.536)	0.137 (0.497)	0.594** (0.272)	1.859*** (0.602)	0.617** (0.292)
Female education	0.010 (0.006)	-0.013 (0.013)	0.002 (0.005)	0.018 (0.015)	-0.001 (0.012)
Centre (ref. North)	0.003 (0.081)	0.026 (0.127)	0.001 (0.063)	0.087 (0.158)	-0.012 (0.078)
South and Islands (ref. North)	-0.046 (0.054)	-0.159 (0.130)	-0.040 (0.050)	0.208** (0.089)	0.024 (0.084)
Taste discrimination reaction	0.988 (1.107)	1.197 (1.064)	1.475** (0.677)	3.112 (1.807)	0.353 (0.864)
Female education	0.011 (0.007)	-0.010 (0.015)	0.003 (0.006)	0.016 (0.022)	-0.004 (0.012)
Centre (ref. North)	-0.005 (0.070)	0.024 (0.126)	-0.013 (0.053)	0.041 (0.175)	-0.028 (0.087)
South and Islands (ref. North)	-0.014 (0.078)	-0.140 (0.132)	0.011 (0.069)	0.356** (0.149)	0.067 (0.088)
Observations	60	60	60	60	60

Notes: Cluster-robust standard errors (region) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All specifications include year fixed effects (omitted from the table). Female education is expressed in percentage points.

Table E10: Regional claim outcomes and beliefs about causes of gender inequalities

E OEQ: Beliefs Policies

Table F1: Categories for beliefs about policies to reduce gender inequalities

ID	Category	Examples from answers	Main keywords
Policy categories			
1	Family support policy	• “Hire more women; provide free daycare; bonuses for babysitters.” • “Promote flexibility for men and women.” (<i>also fits Flexible work</i>) • “Provide workplace facilities like daycare centers.” • “Social bonuses for newborns; enable smart working for pregnant women.”	Daycare; babysitters; bonuses; flexibility; work-life balance; smart working; newborns; pregnancy; family; parental leave; paternity; maternity.
2	Pro-quota	• “Equal percentage between men and women, 50% in each company.” • “Put 50% women in senior management positions.” • “Require by law a minimum number of women in management.”	Equality; 50/50; quotas; mandated share; legal requirement.

Table continues on the next page

ID	Category	Examples from answers	Main keywords
3	Anti-quota	• “Eliminate gender quotas; women have the same abilities—this should be understood.”	Eliminate quotas; opposing quotas.
4	Anti-discrimination policy	• “Oversight of contracts by INPS and the Ministry of Labor.” • “Equalize salaries for equal positions regardless of gender.” • “Blind recruitment; remove gender from applications.”	Equality; oversight; audits; equal pay; blind hiring; inform; law; enforce.
5	Incentives (non-quota)	• “Government should incentivize firms so women can access managerial roles.” • “Signal that salaries must be equal; otherwise higher taxes for noncompliant firms.”	Incentives; tax credits/penalties; bonuses.

Table continues on the next page

ID	Category	Examples from answers	Main keywords
6	Cultural change	• “Civic education in schools teaching gender equality (e.g., chores taught to all), removal of stereotypes.” • “Structure education on equal opportunities; devalue sexist culture.”	School; education; media; stereotypes; culture change.
7	Merit	• “Ensure fair salaries reflecting true abilities, regardless of gender.” • “Pay and promotions based on skills, not gender.” • “Evaluate competencies irrespective of gender.”	Merit; skills; competence; performance; meritocracy.
8	No idea	• “I don’t know.” “No idea.”	I don’t know; no idea.
9	Not classifiable	—	—

	(1) Demad	(2) Supply	(3) Info Causes	(4) NoInfo	(5) Total	(6) p(D=S)	(7) p(D=IC)	(8) p(D=C)	(9) p(S=IC)	(10) p(S=C)	(11) p(IC=C)
Should the government do more to decrease gender inequalities?											
No	52	91	48	52	243	0.002	0.892	0.299	0.001	0.052	0.248
Observations	629	661	596	518	2404						
Yes											
Family Support Policy	131	144	141	119	535	0.675	0.234	0.381	0.429	0.627	0.787
Pro-Quota	73	45	62	36	216	0.416	0.969	0.364	0.413	0.896	0.135
Anti-Quota	10	6	7	7	30	0.817	0.688	0.601	0.889	0.497	0.393
Against Discrimination Policy	398	341	351	276	1366	0.000	0.116	0.001	0.009	0.627	0.564
Incentives	103	99	116	89	407	0.159	0.716	0.035	0.035	0.305	0.327
Cultural Change	213	191	195	154	753	0.055	0.671	0.135	0.143	0.755	0.284
Merit	98	110	96	77	381	0.605	0.801	0.738	0.799	0.408	0.568
No Idea	19	45	26	30	120	0.002	0.212	0.021	0.061	0.478	0.276
Observations	629	661	595	518	2403						

Note that one answer can be classified with more than one category so the number of observation refer just to the total answers

Table F2: Policies to decrease inequalities: Number of observations and balance test, p-value from t-tests

	Family Support Policy	Pro-Quota	Anti-Quota	Against Discrimination Policy	Incentives	Cultural Change	Merit
Family Support Policy	1.000						
Pro-Quota	-0.238*	1.000					
Anti-Quota	0.093*	-0.467*	1.000				
Against Discrimination Policy	-0.075*	-0.104*	0.110*	1.000			
Incentives	0.140*	-0.146*	0.024	-0.036*	1.000		
Cultural Change	0.018*	-0.167*	0.176*	-0.102*	-0.066*	1.000	
Merit	-0.076*	-0.160*	0.174*	0.342*	-0.050*	0.142*	1.000

* $p < 0.05$

Table F3: Policies to decrease inequalities: Correlation matrix

	Men	Women	t-test	p-value
Family Support Policy	0.158	0.275	-6.945	0.000
Pro-Quota	0.661	0.688	-0.498	0.619
Anti-Quota	0.113	0.081	0.098	0.334
Against Discrimination Policy	0.544	0.588	-2.134	0.033
Incentives	0.140	0.194	-3.526	0.000
Cultural Change	0.302	0.322	-1.053	0.293
Merit	0.148	0.167	-1.289	0.198
No Idea	0.067	0.036	3.583	0.000
N	2402			

Table F4: Categories Policies to decrease inequalities: means, t-tests and p-value by gender

F OEQ: Motivation for supporting/opposing Quotas

Table G1: Categories for motivations supporting gender quotas

ID	Category	Examples from answers	Main keywords
Motivations (Yes = quotas are useful)			
1	Fairness / equality	• “There shouldn’t be gender differences.” • “It would put men and women on equal footing.” • “It’s right for men and women to be equal.”	Right; equality; equal footing; nondiscrimination; fairness; parity.
2	Signaling / information value	• “Signals that commitment to equality is essential to start a working group.” • “Lets us assess a woman’s qualifications and whether she is up to the task.” • “By providing information, it helps dissolve prejudices about managers.”	Information; signaling; skills; merit; quality; prejudice reduction.

Table continues on the next page

ID	Category	Examples from answers	Main keywords
3	For women (supply-side)	• “Women would feel less anxious, knowing opportunities are equal.” • “Some women feel guilty about prioritizing work over caregiving.” • “Quotas would encourage women to be more competitive.”	Encourage; confidence; empowerment; competitiveness; role models.
4	No specific justification	• “Personal reasons lead me to think it could work.” • “No specific reason; it’s just my view and I find it right.”	Personal view; unspecified reason; intuition.
5	Explicit mention of video	• “For the reasons outlined in the video I watched.”	Video; treatment content.

Table G2: Categories for motivations opposing gender quotas

ID	Category	Examples from answers	Main keywords
Motivations (No = quotas are not useful)			
1	Not the right instrument / not effective	• “I don’t believe it can work without incentives.” • “If women aren’t supported during key career stages, quotas alone won’t solve it (e.g., more support for mothers).” • “It seems utopian: I like the concept, but it’s hard to achieve.”	Effectiveness; instrument; support policies; incentives; feasibility.
2	Reverse discrimination / meritocracy	• “This creates the opposite disparity; equality is equality only if applied correctly.” • “It doesn’t respect meritocracy.” • “It doesn’t reward merit but imposes presence.”	Meritocracy; discrimination; disparity; fairness; equality.

Table continues on the next page

ID	Category	Examples from answers	Main keywords
3	Unneeded / backlash	• “Women are culturally equal; issues arise later (e.g., pregnancy leaves long absences and firms suffer).” • “Women don’t need to prove anything.” • “Women don’t need quotas to express professionalism.”	Unnecessary; unintended effects; backlash; professionalism.
4	No specific justification	• “It’s how I feel based on my experience.”	Personal view; unspecified reason.
5	Explicit mention of video	• “For the reasons outlined in the video I watched.”	Video; treatment content.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Demad	Supply	Info Causes	NoInfo	Total	p(D=S)	p(D=IC)	p(D=C)	p(S=IC)	p(S=C)	p(IC=C)
Do you think a quota system could be a useful tool for reducing inequalities?											
Yes											
Fairness	194	271	286	194	950	0.105	0.000	0.006	0.005	0.217	0.157
Signalling	214	165	143	78	600	0.000	0.000	0.000	0.690	0.000	0.000
For women's limits	49	49	105	58	261	0.000	0.536	0.890	0.000	0.001	0.467
Do not justify	21	34	29	20	104	0.108	0.177	0.635	0.822	0.296	0.415
Explicit mention to video	5	3	2	0	10	0.538	0.343	0.066	0.719	0.142	0.213
Observations	482	451	417	324	1674						
No											
Wrong Instrument	69	92	71	78	310	0.109	0.605	0.039	0.290	0.580	0.124
Reverse Discrimination	92	128	118	110	448	0.024	0.016	0.003	0.846	0.427	0.554
Unneeded and Backlash	26	25	40	18	109	0.746	0.046	0.564	0.019	0.780	0.015
Do not justify	14	23	11	25	73	0.178	0.639	0.016	0.075	0.246	0.005
Explicit mention to video	0	0	0	0	0	0					
Observations	147	210	179	194	730						

One answer can be classified with more than one category, the number of observation refer just to the total answers
The number of obs refers to the total number of respondents answering respectively YES or NO to the question:
Do you think a quota system could be a useful tool for reducing inequalities?

Table G3: Motivation for supporting/opposing gender quotas: Number of observations, p-value from t-tests

	Number of motivation mentioned		
	(1)	(2)	(3)
Info Causes	0.049 (0.050)	0.010 (0.063)	0.113 (0.081)
Supply	0.184*** (0.048)	0.258*** (0.062)	0.057 (0.076)
Demand	0.184*** (0.053)	0.212*** (0.065)	0.167* (0.094)
Cultural Causes	-0.045 (0.038)	-0.094** (0.047)	0.026 (0.064)
Female	0.028 (0.038)	0.089* (0.046)	-0.052 (0.061)
Observations	2,370	1,652	718
Sample	Full sample	QuotaUseful = 1	QuotaUseful = 0

Notes: The dependent variable is the number of distinct motivations mentioned in the open-ended justification for supporting/opposing quotas. All specifications include treatment dummies, demographics, job characteristics, and policy attitudes. Justification to the question: *Do you think a quota system could be a useful tool for reducing inequalities?* - Why Yes/No

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, Robust standard errors in parentheses.

Table G4: Effect of Information Treatments on the number of motivations mentioned

	Number of motivation mentioned		
	(1)	(2)	(3)
Info Causes	0.043 (0.057)	-0.009 (0.072)	0.114 (0.096)
Supply	0.193*** (0.057)	0.286*** (0.074)	0.035 (0.093)
Demand	0.180*** (0.061)	0.203*** (0.075)	0.161 (0.112)
Cultural Causes	-0.046 (0.073)	-0.096 (0.100)	0.005 (0.108)
Info Causes $\times$ Cultural Causes	0.023 (0.106)	0.073 (0.135)	-0.010 (0.169)
Supply $\times$ Cultural Causes	-0.028 (0.102)	-0.086 (0.136)	0.073 (0.158)
Demand $\times$ Cultural Causes	0.014 (0.105)	0.031 (0.133)	0.021 (0.188)
Observations	2,370	1,652	718
Sample	Full sample	QuotaUseful = 1	QuotaUseful = 0

Notes: OLS estimates with robust standard errors in parentheses. All models include the full set of demographic, job-related, and policy controls Treatment indicators are interacted with the *Cultural Causes* indicator (generic belief).

The dependent variable is the number of distinct motivations mentioned in the open-ended justification for supporting/opposing quotas.

Justification to the question: *Do you think a quota system could be a useful tool for reducing inequalities?* - Why Yes/No

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, Robust standard errors in parentheses.

Table G5: Effects of Information Treatments on the Number of Motives Mentioned: Heterogeneity by Cultural Causes

	Number of words used in the open-ended response		
	(1)	(2)	(3)
Info Causes	0.049 (0.701)	0.263 (0.776)	0.298 (1.640)
Supply	0.461 (0.688)	1.022 (0.814)	-0.478 (1.473)
Demand	3.458*** (0.908)	3.008*** (0.839)	6.353** (2.735)
Cultural Causes	-2.853*** (0.591)	-3.332*** (0.508)	-2.921 (1.832)
Female	0.859 (0.639)	1.285** (0.572)	1.026 (1.729)
Observations	2,370	1,652	718
Sample	Full sample	QuotaUseful = 1	QuotaUseful = 0

Notes: The dependent variable measures the length (in words) of the open-ended justification provided in response to the question “Do you think a quota system could be a useful tool for reducing inequalities?”, conditional on whether the respondent answered yes or no. All specifications include demographics, job characteristics, and policy attitudes.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, Robust standard errors in parentheses.

Table G6: Effects of Information Treatments on the Length of Open-Ended Justifications

	Number of words used in the open-ended response		
	(1)	(2)	(3)
Info Causes	0.301 (0.911)	0.577 (0.949)	0.082 (2.209)
Supply	0.321 (0.861)	1.007 (1.032)	-0.678 (1.732)
Demand	3.318*** (1.086)	3.128*** (1.041)	5.452* (2.982)
Cultural Causes	-2.875*** (1.000)	-2.853** (1.229)	-3.864** (1.870)
Info Causes $\times$ Cultural Causes	-0.979 (1.547)	-1.262 (1.606)	0.728 (3.297)
Supply $\times$ Cultural Causes	0.449 (1.326)	-0.075 (1.537)	0.535 (2.780)
Demand $\times$ Cultural Causes	0.496 (1.685)	-0.512 (1.619)	3.174 (5.009)
Observations	2,370	1,652	718
Sample	Full sample	QuotaUseful=1	QuotaUseful=0

Notes: The dependent variable measures the length (in words) of the open-ended justification provided in response to the question “Do you think a quota system could be a useful tool for reducing inequalities?”, conditional on whether the respondent answered yes or no. All specifications include demographics, job characteristics, and policy attitudes. .

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, Robust standard errors in parentheses.

Table G7: Effects of Information Treatments on the Length of Open-Ended Justifications: Heterogeneity by Cultural Causes

G Complete Survey

Table I1: Survey sections, topics, and main questions

Section	Topics	Main Questions
1	Background; job & firm characteristics	• Year/country of birth; place of residence; marital status; # children; education. • Firm size (employees); sector; job role.
1	Importance of gender equality	<i>Rank</i> six challenges for the next five years (1 = most important): innovation; productivity; diversity & gender equality; technological progress; teamwork; environmental impact.
2	Incentivized beliefs on inequality (facts)	• For managers (age 30–49, FT, firms >10 employees): “For every €100 earned by a comparable man, how much does a woman earn?” • “What % of CEOs are women among the 50 largest listed firms in Italy?”
3	Beliefs on causes of inequality	• “What are the causes of gender inequalities in pay and managerial positions?”
3	Beliefs on policies	• “Should the government do more to promote gender equality in managerial positions? (Yes/No)” • <i>If Yes</i>: “Which policy would be effective?” • Brief description of gender quotas; “Were you already aware of quotas? (Yes/No)”

Table continues on the next page

Section	Topics	Main Questions
4	Information treatments	Assignment to <i>Demand</i> , <i>Supply</i> , or <i>Info Causes</i> (or <i>No Info</i>).
5	Self-reported policy demand	• “Some argue quotas are justified; others say they unfairly advantage women. Are quotas a useful tool to reduce inequalities? (Yes/No) Why?” • <i>If Yes</i>: “Would you agree to introduce quotas in your workplace? (Yes/No)”
5	Secondary outcomes (perceptions/mechanisms)	• If a quota (more women/fewer men) were introduced, to what extent would it: make people work less; apply for higher positions; stop teamwork? • Who is advantaged/disadvantaged: working women/men; non-working women/men? • “How important is it for men(women) to act against unfair dynamics?” (4-point scale) • Agreement (1 “Strongly agree”–5 “Strongly disagree”): – “If women wanted top positions, they should work harder.” – “Past norms/biases hindered women’s self-determination and access to top jobs.” • Policy importance (4-point scale): gender-sensitive language; flexible hours; paternity leave; pay transparency.

Table continues on the next page

Section	Topics	Main Questions
6	Donation (incentivized behavior)	• Lottery: €500 voucher prize. Before knowing outcome, option to donate €0–€500 to the European Women’s Lobby (50% match by researchers). • “Do you wish to donate part of potential winnings?” (Yes/No) → “Choose amount.”
7	Final questions & feedback	• “Is gender important for career prospects in your workplace? (Yes/No)” • “Are there gender-equality programs in your organization? (Yes/No)” <i>If Yes & female</i>: “Have you benefited?” • UNI PdR 125:2022 certification: heard of it? (Yes/No) <i>If Yes</i>: has your org undertaken it? (Yes/No/Don’t know) • Perceived political bias of the study: left-leaning / right-leaning / none. • <i>If treated video</i>: “Would you like more info/sources used in the video? (Yes/No)” • Incentive follow-up (beliefs task): if selected for €5/answer, choose receive voucher or donate to: Save the Children; Doctors Without Borders; Emergency; ActionAid. • Provide correct answers for the two incentivized belief questions (wage ratio; CEO share).

Table I1: Description of the survey

Young Adults and Gender Based Violence: The Role of Social Norms and Controlling Behaviors

Federica Esposito* Elisa Orlandi†

February 16, 2026

Abstract

Misperceptions of social norms can sustain tolerance for coercive and controlling behaviours in intimate relationships. Using a preregistered online experiment with 3,900 young adults from Italy, Sweden and Romania, the study tests whether providing accurate information about peers' attitudes shifts beliefs and policy preferences related to violence against women. Participants systematically overestimated peer acceptance of controlling behaviours. Corrective feedback reduced support for coercive options, particularly among men and those with larger initial biases. Effects on policy preferences were asymmetric—strengthening support for punitive measures while reducing perceived need for preventive or support-based interventions. Behavioural effects are limited but consistent with a shift toward greater willingness to support anti-violence initiatives.

*Department of Economics, University of Bologna. federica.esposito@unibo.it

†Department of Economics, University of Salerno. elisa.orlandi@unisa.it

1 Introduction

The European Commission defines gender-based violence as “violence directed against a person because of that person’s gender or violence that affects persons of a particular gender disproportionately,” encompassing physical, sexual, psychological and economic harm, including technology-facilitated abuse (e.g., non-consensual sharing of intimate images, cyberstalking) (European Commission, 2025). Controlling behaviours, such as monitoring a partner’s activity, restricting social interactions, or demanding access to personal devices, might fall under the psychological forms covered by this definition. GBV imposes large private and public costs through poorer health, lost earnings, and expenditures in healthcare, policing and justice (European Institute for Gender Equality, 2021). It is therefore a central policy concern in the European Union. The EU response combines criminalisation and victim protection with preventive measures such as awareness campaigns, school-based education and professional training (European Commission, 2020; Council of Europe, 2011). In this study we focus on one of these preventive levers: providing concise, credible information about peers’ attitudes about controlling behaviours, we test whether this information shifts young adults’ own recommendations and policy-relevant choices.

A wide literature shows that social-norm informational treatments can shift meaningful real-world behaviours. Peer-comparison messages have been shown to reduce household energy use (Allcott, 2011; Ayres et al., 2013)¹, decrease residential water consumption (Ferraro and Price, 2013; Bernedo et al., 2014), strengthen pro-environmental practices such as hotel towel reuse (Goldstein et al., 2008), raise charitable contributions (Shang and Croson, 2009) and increase tax compliance (Hallsworth et al., 2017).

Building on this evidence, the central mechanism we are interested in is informational: people form beliefs about what they think their peers would recommend in a given situation, and those beliefs can be wrong. In other words, individuals hold second-order beliefs (perceived social norms) about peers’ likely recommendations; these beliefs often deviate from the realised peer averages, creating a misperception gap between perceived and actual

¹For energy use, several studies document that the effectiveness and magnitude of norm-based behavioural changes can vary substantially across contexts and with baseline levels of household consumption (Costa and Kahn, 2013; Andor et al., 2020; Schultz et al., 2007).

social norms. Such gaps can distort personal recommendations and related choices toward what people believe others endorse, rather than what is actually endorsed (Bicchieri, 2016; Tankard and Paluck, 2016). In many settings, individuals overestimate tolerance for problematic behaviours, which can help sustain harmful norms (Bursztyn and Yang, 2022). At the same time, concise, credible peer information that explicitly corrects (mis)perceived social norms has been shown to shift beliefs and behaviour in domains such as climate donations and policy support (Andre et al., 2024), recycling take-up (Fuhrmann-Riebel et al., 2024), health-related behaviours including alcohol use and vaccination uptake (Neighbors et al., 2004; Moehring et al., 2023), political attitudes (for instance reducing support for partisan violence when misperceptions of rivals’ views are corrected, see Voelkel et al. (2024); Mernyk et al. (2022)) and finally attitudes toward gender roles and gender-based violence (Bursztyn et al., 2020; Cortés et al., 2024; Bermek et al., 2023; Matavelli, 2024).

Our study aims to contribute to this latter gender-focused stream. In terms of design, it is close and takes inspiration from Cortés et al. (2024). Like them, we use short vignettes to elicit both personal recommendations and second-order beliefs about peers’ recommendations and randomise whether participants receive peer information (actual social norm).² We adapt this structure to the GBV domain and to a population of EU young adults, focusing on everyday controlling behaviours rather than mothers’ employment. In terms of topic, we are close to Bermek et al. (2023), who study victim-blaming around severe intimate-partner violence in Turkey. The focus and mechanism, however, differ in three respects: first, we study early-stage forms of psychological GBV, control and jealousy, rather than severe IPV; second, we implement a strict misperception-correction tied to measured gaps in perceived peer views, whereas they rely on norm salience and dynamic feedback; third, we have a different sample, we target European Gen Z.

Focusing on everyday control and jealousy is policy-relevant for two reasons. First, EU frameworks explicitly classify coercive and controlling behaviours as forms of psychological violence (European Institute for Gender Equality, n.d.a,n) and recent EU legislation recog-

²Our design follows the “vignettes + second-order beliefs” approach in Cortés et al. (2024), but this is only one way to elicit social norms. Alternative approaches include incentivized coordination games à la Krupka and Weber (2013) and the procedure in Bicchieri and Xiao (2009) that separates empirical from normative expectations. It is important to note that different methods can yield different estimates of social norms (Barigozzi and Montinari, 2025).

nises coercive control within domestic violence ([European Parliament and Council, 2024](#)). Second, controlling/coercive behaviour is consistently associated with greater risk of more severe abuse: in a WHO multi-country study, women with more controlling partners were more likely to experience physical or sexual IPV ([Garcia-Moreno et al., 2006](#)) and longitudinal evidence indicates that baseline coercive control predicts subsequent physical IPV ([Hilton and Radatz, 2024](#)).

Motivated by this rationale, we study whether providing information about peers’ attitudes to correct misperception in social norms shifts personal recommendations in situations involving control, jealousy and psychological forms of GBV. We implement a preregistered, between-subjects online experiment with 3,915 participants aged 18–25 in Italy, Sweden and Romania. These countries were chosen because they span a broad range across the EU Gender Equality Index ([European Institute for Gender Equality, 2024](#)), providing meaningful variation in baseline attitudes. Participants evaluate four short vignettes that depict common situations among young adults. For each vignette, we first elicit second-order beliefs about peers’ recommendations and then participants’ own recommendations (first order beliefs). The treatment provides normative feedback reporting the realised average recommendation of same-country, same-gender peers in our study. The control group receives no feedback. We incentivise accuracy in second-order beliefs with a monetary prediction bonus. Because baseline norms differ substantially across countries in our sample, we also examine heterogeneity in both beliefs and treatment effects across Italy, Sweden and Romania.

Outcomes extend beyond vignette judgments. We measure support for GBV prevention policies that include funding for support services, school programmes, penalties and awareness campaigns. We also elicit an incentivized donation choice to an anti-violence NGO. We ask respondents to assess the adequacy of media coverage of femicide. We also measure demand for information by offering the option to access additional material on GBV. The preregistration specifies average treatment effects on first-order recommendations, effects on these downstream outcomes, and heterogeneity by gender and by the size of the initial misperception gap.

Identification rests on three features. First, the feedback is locally credible because it reports realised peer averages matched on country and gender within the same study, not

generic benchmarks. Second, assignment is between subjects, so exposure to feedback is exogenous at the individual level. Third, we collect covariates, including recent exposure to GBV or femicide coverage, that enter robustness checks to limit concerns that contemporaneous shocks rather than information drive results.

We find that participants systematically overestimate their peers’ acceptance of coercive and controlling behaviours, perceiving social norms as more conservative than they actually are. Correcting these misperceptions significantly reduces support for restrictive options, particularly among those who initially overestimated peer tolerance. The attitudinal effects are heterogeneous: men update their personal beliefs downward, while women, with more progressive baselines, show limited or even compensatory shifts. The treatment’s impact on policy preferences varies across domains and vignettes. For instance, correcting overestimation increases support for punitive measures but can lower demand for preventive or support-based policies, indicating that perceived social prevalence shapes perceived policy urgency. Behaviourally, effects are modest: while nearly all respondents donated to an anti-violence NGO, the treatment only marginally increased donations among those who had overestimated peer conservatism. Overall, the results show that social-norm feedback acts as a targeted corrective mechanism, shaping values and policy support in gendered and context-dependent ways.

The paper contributes to the economics of social norms and information design in three ways. It isolates the causal effect of belief correction using realised, locally matched peer information in sensitive and policy-relevant domains, namely control and jealousy, which, to the best of our knowledge, have not yet been analysed in this literature. It focuses on early adulthood and compares heterogeneous European contexts. It links norm updating to policy preferences, an incentivised monetary decision and demand for information, and studies heterogeneity by gender and by the magnitude of initial misperceptions.

The rest of the paper is organized as follows. Section 2 describes the experimental design, and Section 3 outlines the main hypotheses. Section 4 presents the data, while Section 5 introduces the results, starting with the descriptive analysis of first- and second-order beliefs regarding controlling behaviours. Section 6 focuses on the causal impact of the information treatment and the correction of young adults’ misperceptions about social norms of control

over women. In particular, the first part examines effects on personal beliefs (first-order beliefs), and the second part analyses the impact on policy preferences and demand for institutional support. The final section concludes.

2 Experimental Design

This study employs a between-subjects, randomized experimental design to assess how social norm feedback about peers' attitudes influences individual beliefs and policy preferences related to gender-based violence (GBV). The sample focuses on young adults aged 18–25 in three EU countries, Italy, Sweden and Romania. The study was preregistered on the Open Science Framework (OSF) at osf.io/53z76 prior to data collection. The experiment was conducted in September 2025 and it included a pilot phase followed by the first and the second wave.

Treatment and Vignettes

The two waves defined, respectively, the control and treatment groups to which the participants were randomly assigned. Both groups were shown four different vignettes describing situations involving young adults, each depicting actions by the vignette characters that could be seen as acts of violence or control towards others. We initially developed six scenarios and then selected four for the main survey. The details on the selection process of the vignettes are reported in Appendix A.

Each vignette aims to test a different aspect of gender-based violence, which are often underestimated but nonetheless important in defining controlling attitudes and risks. Three out of four vignettes focus on controlling behavior in relationships, which can relate to study or job opportunities, everyday life, or social media. The last vignette tests another important aspect: violence related to sharing material without consent—a highly relevant topic in the social media era, especially among young adults. The order of vignettes was randomized.

The vignettes can be summarized as follows³:

³The full text is available in Appendix A.

- **Study Abroad:** A woman has been accepted to study abroad, but her partner is worried about the consequences of this period apart. Option A suggests the partner communicates discomfort by implying she should not leave, while Option B shows a more supportive approach where he expresses concerns but does not stop her.
- **Share Location:** A man is unsure about his long-distance relationship. Option A has him asking his partner to share her location to know what she is doing, while Option B involves organizing a nightly video call to catch up on her life.
- **Social Media:** A woman posts a bikini photo on social media, and her boyfriend feels uncomfortable about the suggestive comments she receives. He asks her either to delete the photo or not to post similar photos again. Both options contain controlling nuances, but unlike other vignettes where one option is more conservative than the other, here both options reflect control to some degree. This allows us to observe what happens when the more appropriate action is missing.
- **Private Picture:** A man is uncomfortable because, during a night out with friends, another man shows a photo of his ex-girlfriend in a bikini and makes comments. He feels this action is wrong and non-consensual. In Option B, he communicates this feeling directly, stating that the behavior is inappropriate, while in Option A, he remains silent.

Each group was asked to state the respondents' second-order beliefs—that is, the percentage of others in their country, age, and gender group who would recommend both Option A and Option B—and their first-order beliefs, meaning the percentage with which they themselves would recommend each possible action.⁴ For each vignette, there were two possible actions. What differed between the two groups was that, after collecting data in the first wave, the treatment group in the second wave received feedback about the social norm. Specifically, for only one of the two options—namely, the option perceived as more problematic or negative—participants were shown the actual percentage of peers (within their group, age, and gender) who recommended that option. In the cases of controlling behaviors, this

⁴The percentages were not required to sum to 100, so each option could be recommended on a scale from 0 to 100%.

was the more coercive option; in the case of sharing without consent, it was the option of remaining silent.

To ensure that participants paid close attention to their second-order belief estimates, we provided a monetary incentive: a €10 bonus was awarded to the 10 participants whose predictions of the peer average for a randomly selected vignette were closest to the actual values. This incentive mechanism aligns with common practices in the literature and helps improve the reliability of belief elicitation. This feedback serves as a real-world social norm signal and allows us to examine how correcting misbeliefs affects individual attitudes.

Outcomes

Following the assessment of second-order beliefs and the provision of social norm feedback in the treatment group, we elicited participants' first-order beliefs concerning their personal valuations of the vignette scenarios. The primary outcome of interest was thus the personal beliefs attributed to each vignette, and how the correction of misperceived social norms regarding the most negative option influenced these personal valuations.

Subsequently, we examined the impact on a secondary key outcome: policy support. Participants were asked to rate, on a scale from 1 to 100, their level of agreement with various policy interventions designed to address gender-based violence. These policies included increasing funding for anti-violence centers, expanding affective education programs in schools, strengthening legal penalties, and promoting public awareness campaigns.

As a behavioral measure, we also assessed participants' willingness to donate a potential lottery prize to a nonprofit and NGO organization focused on combating violence against women. This provided an economically meaningful outcome measure, allowing us to capture tangible support for concrete interventions in line with existing economic literature.

As a final outcome, we also assessed participants' willingness to deepen their knowledge about gender-based violence (GBV). Specifically, respondents were given the option to access additional information provided by the European Union's dedicated page on GBV. This allowed us to measure whether the treatment had an effect on participants' willingness to seek further information on the topic.

Randomization and Timing

Participants were recruited through a professional panel provider and assigned to experimental waves as follows:

- **Pilot**
- **Wave 1** : Control group
- **Wave 2** : Treatment group

We completed the entire data collection in September 2025 and began Wave 2 as soon as possible after Wave 1 ended, with an interval of only five days. This short gap between waves, together with a control for recent media exposure to femicide or GBV, reduces concerns about calendar-time effects.

3 Hypotheses

Building on the experimental design (Section 2), we test whether brief, credible information about peers’ attitudes, delivered as social-norm feedback, corrects misperceptions and shifts personal recommendations about the behaviors depicted in the vignettes. We then examine whether these attitudinal shifts extend to related outcomes, including policy support, donation behavior, perceptions of media coverage, and interest in learning more about GBV. Finally, we assess heterogeneity by gender and by the size of individuals’ initial misperceptions. We preregistered the following hypotheses.

H1. Normative feedback shifts personal recommendations. Participants who receive social-norm feedback will report different first-order beliefs (their own recommended action in each vignette) than participants who do not receive feedback. Directionally, we expect lower endorsement of the more problematic option when that option is the one highlighted by the feedback, and a corresponding relative increase for the alternative.

H1.b. Larger effect on the option receiving feedback. Within a vignette, the absolute change in first-order beliefs will be larger for the option that receives feedback than for the option that does not.

H2. Downstream attitudes and behaviors. Relative to Control, participants in Treatment will:

1. express higher support for GBV-prevention policies (funding for anti-violence centers; school programs; harsher penalties; public-awareness campaigns);
2. be more likely to donate any lottery winnings and/or donate a higher amount to an anti-violence NGO;
3. be more likely to judge media coverage of femicide as insufficient;
4. be more willing to deepen their knowledge about GBV when offered the option.

H3. Heterogeneous effects by gender. Treatment effects in H1 and H2 will differ between men and women.

H4. Heterogeneous effects by misperception. Treatment effects in H1 and H2 will be larger for participants who initially misperceive norms by a greater margin (i.e., whose perceived peer support differs more from the observed peer average).

4 Data

Data were collected via Qualtrics through the professional provider VVA. Fieldwork began with a pilot on 1 September 2025 and concluded with the end of Wave 2 on 25 September 2025. In line with the preregistration (osf.io/53z76), we first ran the pilot and then the two experimental waves (Wave 1 = Control; Wave 2 = Treatment). As pre-specified, we applied standard quality screens: we excluded respondents who completed the survey implausibly fast (less than 60 seconds) or took unusually long (more than 3 hours), and we removed respondents who failed either of two embedded attention checks.

The pilot yielded 405 respondents (Italy: 130; Romania: 135; Sweden: 140); pilot observations are excluded from the main analyses. The main experimental sample comprises 3,510 respondents: 1,560 in Wave 1 (Control) and 1,950 in Wave 2 (Treatment). Country×gender quotas were balanced by design. Table 1 reports counts by country, gender and treatment.

Table 1: Participants by Country, Gender and Treatment Status (Main Sample)

Country	Gender	Control (Wave 1)	Treatment (Wave 2)	Total
Italy	Men	260	325	585
	Women	260	325	585
Romania	Men	260	325	585
	Women	260	325	585
Sweden	Men	260	325	585
	Women	260	325	585
Total		1,560	1,950	3,510

Notes. Pilot respondents (N=405) excluded.

Country×gender quotas balanced by design (585 men and 585 women per country).

4.1 Background Characteristics

The analysis sample consists of young adults aged 18–25 residing in Italy, Sweden, and Romania, with country×gender quotas balanced by design. Table 2 reports summary statistics on key background characteristics. Table B1 reports chi-squared tests assessing balance between the treatment and control groups across the main background characteristics.⁵

The results indicate no systematic differences between the treatment and control groups for almost all variables. Got it. Here is a corrected and coherent set of lines that matches your purpose (attriters in treatment vs. control):

To assess whether attrition is differential across experimental arms, we restrict the sample to respondents who drop out and compare their baseline characteristics between Wave 1 (control) and Wave 2 (treatment). Table B2 reports chi-squared tests between treatment status and a set of key background characteristics within the attriting sample. The results indicate no systematic differences across waves for gender, education, religiosity, employment, relationship status, or parental background. The only imbalance concerns country

⁵See Appendix B for the full set of descriptive statistics and balance checks.

composition ($p = 0.011$), which we address by including country fixed effects (and related controls) in all main specifications.

	Control (Wave 1)	Treatment (Wave 2)	Total
Enrolled in or attended university	605	771	1,376
Student	562	737	1,299
Religious	762	968	1,730
In a relationship	923	1,197	2,120
Parents' Characteristics			
Father attended University	442	605	1,047
Mother attended University	566	674	1,240
Mother works	1,115	1,434	2,549
Father works	1,116	1,470	2,586
Geographical Area			
Medium city	495	623	1,118
Metropolitan area	481	575	1,056
Other/unspecified	3	3	6
Rural area	184	241	425
Small city	397	508	905
Observations	1,560	1,950	3,510

Table 2: Main sample characteristics by treatment status.

5 Beliefs About Controlling Behaviour

5.1 First and Second Order Beliefs

Using control-group data, Table B3 reports averages for first-order beliefs (FOB: “what you would advise”) and second-order beliefs (SOB: “share of your peers who would advise”) across the four vignettes. The descriptive statistics reveal a systematic perception gap. In each scenario, respondents tend to believe that peers are more supportive of controlling behaviors (Option A) than they themselves report (gaps of about +5 to +7 percentage points), and at the same time they underestimate support for progressive options (Option B), with negative gaps of similar magnitude (around −5 p.p.). Figure 1 shows that misperceptions are approximately unimodal and bell-shaped, centered close to zero (and slightly to the right in most panels). This indicates substantial heterogeneity: while many respondents are close to accurate (values near zero), the average misperception is positive and the distribution

exhibits a longer right tail, consistent with a subset markedly overestimating peers’ support for the coercive option. Overall, these results indicate a consistent bias: respondents perceive their peers as more conservative than they actually are. In the social norms literature, this pattern is often described as *pluralistic ignorance* (Bicchieri, 2016), which may contribute to the persistence of *bad* norms. In this context, our treatment can help correct these misperceptions and reduce the gap between perceived and actual norms.

Country comparisons highlight systematic cross-national differences. Italians consistently report more progressive first-order beliefs than both Romanians and Swedes: with the exception of the *Private Picture* scenario, support for the coercive option is 6 to 12 percentage points lower with respect to Romania and Sweden. The support for the progressive option shows a less uniform trend: significant differences exist, but they vary across countries and vignettes. Between Romania and Sweden, the patterns are more nuanced: Romanians are generally less conservative than Swedes, showing lower support for the coercive option and higher support for the progressive option, with differences ranging from 1 to 9, although some contrasts do not reach statistical significance.

For second-order beliefs, Romanians and Swedes also tend to perceive greater peer conservatism than Italians: their misperception gap is significantly higher for the coercive option in most vignettes.

These results show that respondents consistently misperceive social norms: they believe peers are more likely to endorse controlling behaviors and less likely to support progressive ones. This “conservatism bias” is strongest when a clearly progressive alternative exists (*Study Abroad*, *Private Picture*, *Share Location*), weaker when the contrast between options is narrower (*Social Media*).

Gender differences in first- and second-order beliefs are large and systematic: across all vignettes, men hold substantially more conservative FOBs and SOBs than women, with gaps of 10–20 percentage points in FOBs and 6–14 points in SOBs ($p < 0.001$ throughout). By contrast, gender differences in the misperception gap (SOB–FOB) are modest and inconsistent. In some scenarios, women show a larger gap (e.g., *Study Abroad A*, *Private Picture A*), while in others men do (most notably *Private Picture B*). The magnitudes of these differences are small (1–5 points) compared to the baseline gender gaps in beliefs, and only a subset

reach statistical significance. Overall, both men and women display the same directional bias of overestimating peer conservatism, even though their baseline expectations differ sharply by gender.

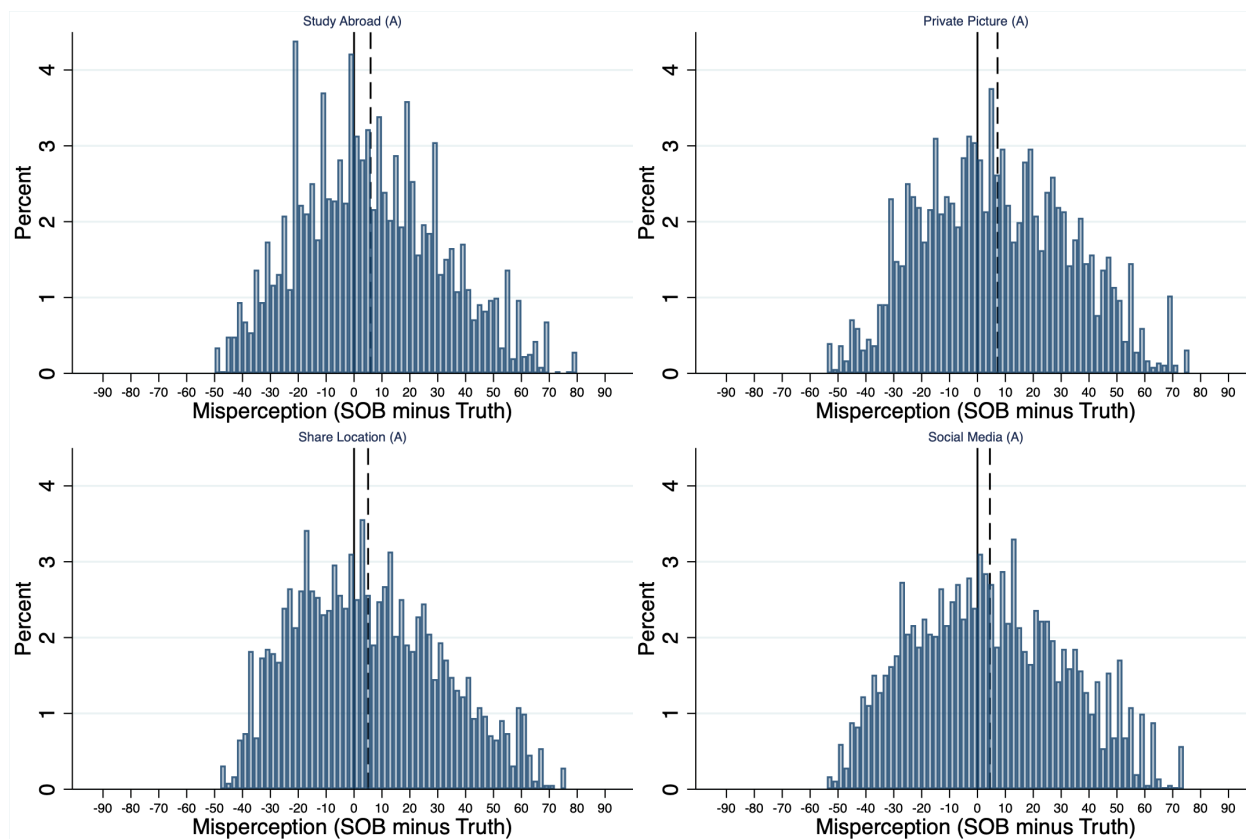


Figure 1: Misperception distribution for the coercive option (SOB minus Truth) across vignettes. *Note:* The solid vertical line indicates 0, and the dashed vertical line indicates the mean.

6 Causal Impact of Correcting Misbeliefs

To assess the causal impact of feedback about actual social norms, we estimated a series of OLS regressions⁶. The first part focuses on estimating the treatment effect on participants' personal beliefs (FOBs) across all vignettes, after they received feedback about the actual social norm and corrected any initial misbeliefs. The second part presents the results related to the demand for institutional support, including specific interventions aimed at reducing violence against women and increasing awareness, as well as a donation to a non-profit NGO focused on the prevention of GBV and support for victims. Before estimating these causal effects, we conducted balance checks and confirmed that there were no significant differences in perceived social norms (second-order beliefs) between treatment and control groups for their perception of the controlling option of the vignettes (see Table B8).

6.1 Personal Beliefs

The empirical results in Tables 3–5 and Figure 3 provide evidence on the effect of the social-norm information on the FOB related to Option A: the more coercive and socially undesirable. The minimal models presented in Tables C2 to C6 do not show any significant average treatment effect.⁷ However, since our hypothesis concerns heterogeneous treatment effects, we now turn to their analysis.

Table 3 reports OLS estimates of the treatment effects on personal beliefs (first-order beliefs) by vignette, including a gender interaction. The interaction coefficients are positive and statistically significant in three of the four vignettes: +3.92 for *Study Abroad* ($p < 0.05$), +3.35 for *Private Picture* ($p < 0.05$), and +2.95 for *Share Location* ($p < 0.05$). These interaction terms indicate that the estimated treatment effect differs by gender in these vignettes. At the same time, the treatment does not shift recommendations in a single direction across scenarios for either gender. The estimated average treatment effects (by gender) are vignette specific. For men, the estimates are small in magnitude in *Private*

⁶See Appendix D for multiple testing adjustments.

⁷Appendix C presents all model specifications, from the minimal version (treatment only, without controls) to the fully specified models including all covariates.

Picture, *Share Location*, and *Social Media*, while they are negative in *Study Abroad*. For women, the implied treatment effects are positive in *Private Picture* and *Share Location* and close to zero in *Study Abroad* and *Social Media*.⁸

In addition, the large and highly significant main coefficients on *Woman* (ranging from -7 to -22 , all $p < 0.01$) confirm substantial gender gaps in baseline attitudes: women are systematically less likely than men to recommend controlling or non-consensual behaviors, in the absence of any feedback.

Taken together, these estimates suggest gender heterogeneity in responses to the information treatment (consistent with Hypothesis H3). For interpretation, it is useful to note that the vignettes refer to men’s behaviour in each scenario, while the norm information is gender specific: men see the distribution of men’s recommendations, while women see the distribution of women’s recommendations about the same scenarios.

We next study heterogeneity by baseline misperceptions. In the next step, we therefore examine treatment effects as a function of respondents’ baseline beliefs and the magnitude (and direction) of their initial misperceptions. Indeed, we expect the impact of information about the actual social norm to vary depending on individuals’ prior beliefs and on how far their perceptions were from the truth at baseline. Tables 4 and 5, together with Figure 3, sheds further light on the mechanisms driving these effects. The vignette-specific regressions in Table 4 show that the treatment interacts strongly with the individual degree of misperception. Across all vignettes, the coefficient on the distance from the actual social norm is large and positive, indicating that individuals who initially overestimated peers’ acceptance of coercive or socially undesirable behavior were also more likely to endorse it themselves. The negative and significant interaction between the information treatment and this distance—especially in the *Study Abroad*, *Share Location*, and *Social Media* vignettes—demonstrates that exposure to the true distribution of beliefs effectively corrected these overestimations, producing an attitudinal shift toward less coercive positions. Similarly, the binary specification using the overestimation dummy confirms that these corrective effects are concentrated among those whose priors were upwardly biased.

⁸For women, the treatment effect equals the sum of the treatment coefficient and the treatment $\times$ woman interaction. The resulting estimates are 1.393 ($p = 0.139$) in *Study Abroad*, 3.057 ($p = 0.003$) in *Private Picture*, 2.501 ($p = 0.011$) in *Share Location*, and 0.446 ($p = 0.675$) in *Social Media*.

The stacked specification in Table 5 pools all vignettes at the individual–vignette level and confirms this interpretation. Here, the negative and highly significant coefficient on the interaction between the information treatment and the continuous distance measure (-0.055 , $p < 0.01$) indicates that the greater the initial misperception, the larger the downward revision in first-order beliefs following feedback. The corresponding interaction with the binary overestimation variable (-4.95 , $p < 0.01$) confirms that corrective responses are strongest among participants who initially believed coercive behaviors to be more socially acceptable than they truly are. Figure 3, derived from the same stacked specification including gender interactions, visually illustrates these heterogeneous effects: for men (blue line), the average marginal effect of the treatment becomes increasingly negative as misperception grows, while for women (orange line) the slope remains nearly flat and indistinguishable from zero. This pattern implies that men—who systematically overestimated the endorsement of coercive norms—respond most to corrective information, whereas women exhibit limited updating. Interestingly, when the distance from the actual social norm is negative—that is, when individuals believed that others were *more progressive* than they actually are—both men and women tend to increase their support for the more coercive or violent options. This suggests that perceiving oneself as more conservative than one’s peers can trigger a form of compensatory adjustment toward behaviors perceived as “normal” within the group, even when such behaviors are socially undesirable. In other words, social-norm information can generate both corrective and reinforcing effects, depending on the direction of the initial misperception. The results confirm that the social-norm feedback operates as a targeted corrective mechanism rather than a uniform shift in beliefs. Among women, estimated effects are close to zero, while among men the feedback meaningfully reduces endorsement of coercive or controlling behaviors. This asymmetric response supports the idea that social-norm information interacts with gender identity and prior beliefs, producing the largest updates where initial misperceptions are highest.

	Study Abroad		Private Picture		Share Location		Social Media	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Info Social Norm	-0.433 (0.905)	-2.522** (1.025)	1.739* (0.913)	-0.288 (1.083)	0.901 (0.931)	-0.444 (0.994)	-0.089 (0.979)	-0.311 (1.115)
Woman	-12.979*** (0.898)	-18.265*** (1.036)	-16.642*** (0.913)	-22.169*** (1.135)	-7.086*** (0.934)	-10.674*** (1.046)	-11.965*** (0.980)	-13.685*** (1.170)
Info Social Norm $\times$ Woman		3.915** (1.390)		3.345** (1.493)		2.945** (1.400)		0.757 (1.543)
Observations	3506	3506	3506	3506	3506	3506	3506	3506
Distance Social Norm	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: The table reports estimated treatment effects (*Info Social Norm*) on the probability (0–100%) of recommending Option A (the more coercive option) across treatment groups. *Distance* measures the deviation between an individual’s perceived and the actual social norm, with positive values indicating overestimation and negative values underestimation. All models include a set of controls: country of residence, religiosity, student status, education level, mother’s education and employment, being in a relationship, and whether the respondent was born abroad.

Table 3: Effect of Information Treatment on personal beliefs by Vignette

	Study Abroad		Private Picture		Share Location		Social Media	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Info Social Norm	-0.225 (0.651)	3.831*** (1.012)	1.381** (0.695)	3.029*** (1.099)	1.411** (0.657)	4.537*** (1.032)	0.488 (0.745)	3.347*** (1.154)
Distance Social Norm	0.691*** (0.022)		0.587*** (0.024)		0.732*** (0.022)		0.695*** (0.022)	
Info Social Norm $\times$ Distance	-0.057* (0.031)		0.000 (0.033)		-0.075** (0.030)		-0.093*** (0.030)	
Positive Distance/Overestimate		31.182*** (1.097)		26.979*** (1.186)		32.546*** (1.124)		30.914*** (1.259)
Info Social Norm $\times$ Positive Distance/Overestimate		-6.218*** (1.476)		-2.722* (1.571)		-5.700*** (1.510)		-5.245*** (1.670)
Distance (abs. value)		0.175*** (0.030)		0.122*** (0.031)		0.224*** (0.031)		0.197*** (0.031)
Observations	3506	3506	3506	3506	3506	3506	3506	3506

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: The table reports estimated treatment effects (*Info Social Norm*) on the probability (0–100%) of recommending Option A (the more coercive option) across treatment groups. *Distance Social Norm* measures the deviation between an individual’s perceived and the actual social norm. *Positive Distance/Overestimate* is a dummy =1 if *Distance Social Norm* is positive, indicating an overestimation of people overestimating their peers’ support for the more coercive option. All models include a set of controls: country of residence, religiosity, student status, education level, mother’s education and employment, being in a relationship, and whether the respondent was born abroad.

Table 4: Effects of Information Treatment and Distance on Personal Beliefs

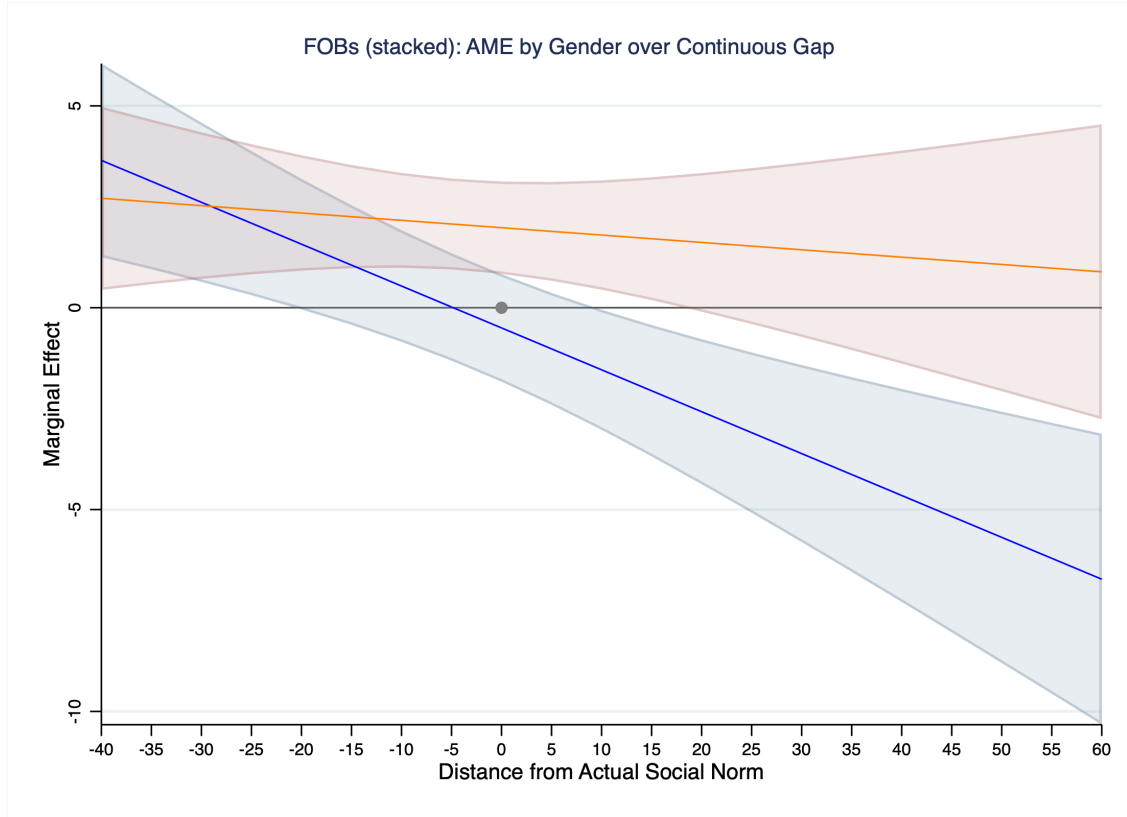


Figure 2: Marginal effects of the treatment on perceived social norms across vignettes by gender (blue line= Man; orange line= Woman). The figure reports average marginal effects of the treatment, with 95% confidence intervals, for each vignette (Study Abroad, Private Picture, Location Sharing, and Social Media).

	Personal Beliefs (FOB)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Info Social Norm	0.505 (0.728)	0.529 (0.672)	-0.453 (0.903)	-1.409 (0.952)	-0.897 (0.702)	0.783* (0.451)	3.675*** (0.695)
Private Picture	1.023** (0.483)	1.023** (0.484)	0.243 (0.725)	1.023** (0.484)	1.592*** (0.424)	1.587*** (0.423)	1.438*** (0.446)
Share Location	2.813*** (0.509)	2.815*** (0.510)	1.644** (0.787)	2.815*** (0.510)	2.006*** (0.450)	2.012*** (0.449)	2.231*** (0.475)
Social Media	5.112*** (0.512)	5.099*** (0.512)	4.864*** (0.781)	5.099*** (0.512)	6.048*** (0.444)	6.042*** (0.444)	5.280*** (0.475)
Info Social Norm $\times$ Private Picture			1.402 (0.973)				
Info Social Norm $\times$ Share Location			2.107** (1.031)				
Info Social Norm $\times$ Social Media			0.422 (1.035)				
Woman		-12.168*** (0.669)	-12.168*** (0.669)	-14.327*** (1.035)	-16.217*** (0.728)	-14.667*** (0.476)	-12.276*** (0.521)
Info Social Norm $\times$ Woman				3.875*** (1.345)	2.733*** (0.959)		
Distance Social Norm					0.640*** (0.010)	0.670*** (0.015)	
Info Social Norm $\times$ Distance						-0.055*** (0.020)	
Positive Distance/Overestimate							30.453*** (0.748)
Info Social Norm $\times$ Positive Distance							-4.945*** (0.991)
Distance Social Norm (abs. value)							0.176*** (0.019)
Observations	14040	14024	14024	14024	14024	14024	14024

Robust standard errors clustered at the individual level in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: The table reports estimated treatment effects (*Info Social Norm*) on the probability (0–100%) of recommending Option A (the more coercive option) across treatment groups. *Distance Social Norm* measures the deviation between an individual’s perceived and the actual social norm. *Positive Distance/Overestimate* is a dummy =1 if *Distance Social Norm* is positive, indicating an overestimation of people overestimating their peers’ support for the more coercive option. Models (2) to (6) include a set of controls: country of residence, religiosity, student status, education level, mother’s education and employment, being in a relationship, and whether the respondent was born abroad.

Table 5: Effects of Information Treatment and Distance on Personal Beliefs across Vignettes

6.2 Demand for Institutional Response

6.2.1 Donation

We have seen in the previous section that being informed about actual social norms affects support for more restrictive options. With heterogeneous treatment based on gender and on the initial misperception of the actual social norm, But overall, correcting misperceptions about others' attitudes reduces the likelihood of supporting more controlling policies if respondents think their peers would support the controlling option more than the actual value.

We now turn to examine whether these attitudinal effects translate into behavioral changes, focusing on a measure of prosocial behavior included in the experiment: donations to the NGO Women Against Violence Europe (WAVE), a non-profit organization committed to preventing and combating gender-based violence.

Participants were automatically entered into a lottery and could choose to donate part of their potential winnings to WAVE. Overall, 94.7% of participants chose to donate at least €1. The average donation was €47.6 (SD = 29.9). Women donated significantly more than men—€49.95 versus €45.20—and this gender difference is statistically significant ($t = -4.72$, $p < 0.001$). Tables 6 and 7 report the treatment effects on the probability of donating and on the donated amount, respectively. In this first analysis, we find no significant treatment effects on donation behavior, suggesting that information about social norms did not translate into changes in actual behavior. The only difference that we find is that women have higher probability to donate and donate higher amount. When considering respondents who overestimated their peers' support for coercive behavior in the *Study Abroad* vignette (i.e., the binary variable *Distance Social Norm_Study Abroad* in Model 6, Table 5), the coefficient for the distance is negative and significant. A test on the linear combination of the distance coefficient and its interaction with the treatment shows that the estimated treatment effect on the probability of donating is positive (0.74) and marginally significant ($p = 0.068$) among those who overestimated their peers' support for the controlling option. This suggests that, for this subgroup, exposure to social-norm information might have slightly increased their willingness to donate.

	Continuous Distance Measures				Binary Distance Measure	
	(1)	(2)	(3)	(4)	(5)	(6)
Info Social Norm	0.122 (0.151)	0.137 (0.154)	0.130 (0.154)	0.122 (0.188)	0.217 (0.172)	0.339 (0.323)
Info Social Norm $\times$ Woman				0.025 (0.328)		
Woman		0.826*** (0.163)	0.881*** (0.163)	0.868*** (0.241)	0.888*** (0.164)	0.868*** (0.163)
Distance Social Norm_Study Abroad			-0.007* (0.004)	-0.007* (0.004)	-0.013** (0.005)	-0.565** (0.261)
Info Social Norm $\times$ Distance Social Norm_Study Abroad					0.011 (0.007)	0.406 (0.362)
Distance Social Norm_Private Picture			-0.005 (0.003)	-0.005 (0.003)	0.003 (0.005)	0.170 (0.243)
Info Social Norm $\times$ Distance Social Norm_Private Picture					-0.016** (0.007)	-0.429 (0.327)
Distance Social Norm_Share Location			0.004 (0.004)	0.004 (0.004)	0.003 (0.005)	0.157 (0.253)
Info Social Norm $\times$ Distance Social Norm_Share Location					0.001 (0.007)	0.093 (0.343)
Distance Social Norm_Social Media			0.000 (0.003)	0.000 (0.003)	0.004 (0.005)	0.098 (0.247)
Info Social Norm $\times$ Distance Social Norm_Social Media					-0.006 (0.006)	-0.481 (0.337)
Distance_Study Abroad (abs. value)						-0.0027 (0.0055)
Distance_Private Picture (abs. value)						-0.0113** (0.0054)
Distance_Share Location (abs. value)						0.0018 (0.0060)
Distance_Social Media (abs. value)						0.0062 (0.0056)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable equals 1 if the respondent donated a positive amount and 0 otherwise. Estimates come from a logit model. The table reports estimated treatment effects (*Info Social Norm*) on the decision to donate to WAVE. *Distance Social Norm_* measures the deviation between an individual's perceived and the actual social norm for each vignette. For column 1-5 is the continuous measure of distance, for column 6 is a dummy equal to 1 when the *Distance from the Social Norm* is positive, indicating individuals who overestimate their peers' endorsement of the more coercive option. Models (2) to (6) include a set of controls: country of residence, religiosity, student status, education level, mother's education and employment, being in a relationship, and whether the respondent was born abroad.

Table 6: Effects of Social Norm Information on the Probability of Donating (Logit)

	Continuous Distance Measures					Binary Distance Measure
	(1)	(2)	(3)	(4)	(5)	(6)
Info Social Norm	0.023 (0.031)	0.024 (0.031)	0.024 (0.031)	0.035 (0.046)	0.024 (0.033)	0.031 (0.061)
Info Social Norm $\times$ Woman				-0.022 (0.062)		
Woman		0.099*** (0.031)	0.093*** (0.031)	0.105** (0.048)	0.093*** (0.031)	0.098*** (0.031)
Distance Social Norm_Study Abroad			0.0004 (0.001)	0.0004 (0.001)	0.0014 (0.001)	0.035 (0.056)
Info Social Norm $\times$ Distance Social Norm_Study Abroad					-0.0019 (0.001)	-0.087 (0.070)
Distance Social Norm_Private Picture			0.0000 (0.001)	0.0000 (0.001)	-0.0011 (0.001)	-0.021 (0.051)
Info Social Norm $\times$ Distance Social Norm_Private Picture					0.0021 (0.001)	0.095 (0.066)
Distance Social Norm_Share Location			0.0014* (0.001)	0.0014* (0.001)	0.0023** (0.001)	0.105* (0.055)
Info Social Norm $\times$ Distance Social Norm_Share Location					-0.0018 (0.001)	-0.079 (0.070)
Distance Social Norm_Social Media			0.0005 (0.001)	0.0005 (0.001)	-0.0003 (0.001)	-0.001 (0.055)
Info Social Norm $\times$ Distance Social Norm_Social Media					0.0013 (0.001)	0.060 (0.069)
Distance_Study Abroad (abs. value)						0.0022** (0.0011)
Distance_Private Picture (abs. value)						-0.0015 (0.0010)
Distance_Share Location (abs. value)						0.0011 (0.0011)
Distance_Social Media (abs. value)						-0.0001 (0.0011)
Observations	3324	3320	3320	3320	3320	3320

Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports estimated treatment effects (*Info Social Norm*) on the amount decided to donate to WAVE (1–100). *Distance Social Norm_* measures the deviation between an individual’s perceived and the actual social norm for each vignette. For column 1–5 is the continuous measure of distance, for column 6 is a dummy equal to 1 when the *Distance from the Social Norm* is positive, indicating individuals who overestimate their peers’ endorsement of the more coercive option. Models (2) to (6) include a set of controls: country of residence, religiosity, student status, education level, mother’s education and employment, being in a relationship, and whether the respondent was born abroad.

Table 7: Effect of Information Treatment on Amount Donated

To better understand the mechanisms behind donation choices, we included an open-ended question asking participants to explain why they decided to donate the amount they did. To classify the responses, we first manually identified the main themes present in the open-ended answers and then classified each response using an LLM.⁹ Table 8 lists the categories, along with the definitions and indicative keywords provided to the LLM for classification. Table 9 reports the frequency of each category. The most common category is *Justice/Ethics*, capturing motivations based on general moral values (e.g., doing the right thing or acting altruistically) rather than a specific, issue-focused reason.

To assess whether the treatment affected respondents’ stated reasons for donating—beyond any effect on the donation decision itself—we estimate OLS regressions using dummy indicators for each category as outcomes. Our main category of interest is *Specific Support*, which includes responses explicitly linking the donation to combating violence against women and actively contributing to solutions. The results are reported in Tables 10. Among donors, the treatment does not appear to affect the probability that respondents mention category *Specific Support*. Column (3) allows the treatment effect to vary by whether respondents exhibit a positive misperception gap (i.e., they overestimate peers’ support) in each vignette. In this case, there is a negative and significant effect for the interaction between treatment and a positive misperception gap in the *Social Media* vignette (-0.070 , $p < 0.01$). This suggests that, among donors who overestimate peers’ support for the coercive option in the Social Media scenario, receiving norm information reduces the likelihood of giving a *Specific Support* type response by roughly 7 percentage points relative to comparable controls. This pattern may reflect a reduced perceived urgency to “do something”: by correcting overly pessimistic beliefs about peers’ attitudes in the Social Media vignette, the treatment could lower the perceived need for immediate action-oriented support, making *Specific Support*-type motivations less salient. This interpretation is consistent with—and will be revisited in—the next section, where we examine whether the same information also affects respondents’ self-reported support for policy interventions.

⁹We used the ChatGPT o3 model via the OpenAI API.

Category	Definition	Indicative keywords
Specific Support	Direct references to violence against women and the need to stop it, protect victims, support them, take action, or contribute to the cause/campaign.	protection; safety; femicide; violence; support; act; need; stop; abuse; control; coercive; rape; harassment; equality; campaign; participate; purpose
Justice Ethics	Ethical or moral motivations grounded in universal values (e.g., fairness, responsibility, solidarity, altruism, charity).	right/fair; justice; equitable; responsibility; duty; help; solidarity; humanity; compassion; kindness; altruism; charity; generosity; good deed; conscience
Personal Experience	Mentions of personal or family experiences related to the topic.	personal experience; family; mother; victim
Distrust	Doubts or mistrust about how donations are used or whether they are effective.	skeptical; doubts; distrust; where it goes; what they do; ineffective
Financial Constraints	References to personal financial hardship or limited ability to donate.	unemployed; student; rent; expenses; financial hardship; cannot afford; salary; poverty
Not Explained	Vague, non-informative, or unclear responses that do not provide an interpretable reason.	no reason; nothing; unclear; "I don't know"
Denial	Denial or minimization of violence against women or the relevance of the issue.	does not exist; not true; not a problem; exaggerated; does not happen

Table 8: Categories, definitions, and indicative keywords used to guide the semantic classification of open-ended responses.

Categories	All		Donation>0	
	Freq.	%	Freq.	%
Denial	32	0.91	18	0.54
Distrust	89	2.54	65	1.96
Financial constraints	556	15.84	471	14.17
Justice / ethics	1,402	39.94	1,396	42.00
Not explained	853	24.30	797	23.98
Personal experience	75	2.14	74	2.23
Specific support	503	14.33	503	15.13
Total	3,510	100.00	3,324	100.00

Table 9: Distribution of reasons in the full sample and among donors

	Reason to Donate: Specific Support		
	(1)	(2)	(3)
Info Social Norm	0.002 (0.013)	0.003 (0.012)	0.008 (0.024)
Positive Distance_Study Abroad			-0.040* (0.021)
Info Social Norm $\times$ Positive Distance_Study Abroad			0.023 (0.027)
Positive Distance_Private Picture			-0.042** (0.020)
Info Social Norm $\times$ Positive Distance_Private Picture			0.007 (0.026)
Positive Distance_Share Location			-0.029 (0.020)
Info Social Norm $\times$ Positive Distance_Share Location			0.029 (0.027)
Positive Distance_Social Media			0.030 (0.020)
Info Social Norm $\times$ Positive Distance_Social Media			-0.070*** (0.027)
Controls	No	Yes	Yes
Observations	3,324	3,320	3,320

Table 10: Treatment effect on the Reasons to Donate: Specific Support

Notes: OLS estimates. The dependent variable is equal to 1 if the respondents suggest *Specific Support* as the main reason for their donation. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls include country of residence, gender, religiosity, university, student status, born abroad, area, mother work, having a partner, and mother's university education.

6.3 Support for Policies to Prevent and Contrast Violence Against Women

As an additional measure of support for institutional intervention, respondents were asked to rate, on a scale from 0 to 100, their level of agreement with a range of policy actions designed to address violence against women. The results summarized in Tables 11–14 examine how exposure to information about peers’ attitudes toward controlling behaviors over women affects support for preventive and corrective policy measures. Across all outcomes—including sexual education in schools, awareness campaigns, harsher punishment for perpetrators, and increased funding for anti-violence centers—the information treatment itself does not exert a significant average effect. However, heterogeneous patterns emerge once we account for individuals’ misperceptions of social norms. These heterogeneous effects are discussed with reference to Model (6) in each table, which includes the binary distance indicators—capturing the direction of the gap between perceived and actual social norm— and their absolute deviations. For *sexual education*, the distance related to the *Study Abroad* vignette is negative and statistically significant ($\beta = -3.18, p < 0.05$), indicating that respondents who overestimate peers’ acceptance of coercive control in the context of women’s decision to study abroad are less likely to support educational initiatives. This suggests that inflated perceptions of social tolerance toward control behaviors may translate into lower perceived need for preventive interventions.

A similar pattern appears for *awareness campaigns*, where greater misperception regarding the *Private Picture* vignette correlates with weaker support ($\beta = -0.035, p < 0.10$). Again, this indicates that respondents who believe coercive behaviors are more widely tolerated show less inclination to endorse awareness-raising measures.

The results for *harsher punishment* in Table 13 highlight a more dynamic interaction. Respondents who overestimated peers’ tolerance toward controlling a partner’s location reported significantly lower support for punitive measures ($\beta = -4.63, p = 0.002$). However, the interaction between the treatment and this misperception is positive and highly significant ($\beta = 5.68, p = 0.004$), suggesting that the informational correction effectively reversed this attitudinal gap. Once exposed to accurate information about their peers’ actual disap-

proval of such behaviors, these individuals expressed markedly higher support for harsher punishment.

Interestingly, the pattern reverses when considering support for *anti-violence centers* (Table 14). In this case, respondents who overestimated peers' tolerance toward controlling behaviors on social media expressed substantially higher support for institutional responses ($\beta = 5.67, p < 0.001$). Yet, the interaction between treatment and this misperception is negative and highly significant ($\beta = -6.32, p = 0.001$), and the linear combination test confirms that treated individuals with positive misperceptions ultimately report lower support ($\beta = -6.89, p = 0.002$). This implies that learning that others are, in fact, less tolerant of such behaviors reduces the perceived urgency of institutional interventions, possibly because the problem is seen as less socially widespread.

The direction and context of misperceptions matter for how individuals respond to normative information. When respondents *overestimate* peers' acceptance of controlling behaviors, informational corrections can either increase or decrease policy support depending on the perceived domain of action. For punitive measures, realizing that others disapprove of coercive control seems to reinforce support for punishment, while in the case of support for anti-violence centers, it may instead reduce the perceived need for institutional involvement. This asymmetry underscores that social norm updating affects not only moral evaluations but also perceptions of institutional necessity, depending on whether the policy targets prevention, punishment, or support services.

Across all policy domains, background covariates display remarkably consistent patterns. Women systematically report higher levels of support across all specifications, with coefficients between 7 and 12 points ($p < 0.001$), confirming gender as the strongest predictor of policy endorsement. Religiosity is negatively associated with support ($\beta \approx -2, p < 0.05$), while respondents whose mothers are employed show significantly greater support ($p < 0.05$). Those in a relationship show less support, particularly for punitive and institutional measures. Finally, respondents in smaller cities or rural areas express weaker support for anti-violence centers, consistent with lower institutional salience in less urban contexts.

Overall, these results highlight that both informational and structural factors shape young adults' support for gender-based violence policies. Misperceptions of peers' norms influence

the perceived need for intervention, while stable sociodemographic patterns—especially gender, religiosity, and family background—shape broader attitudinal divides in how institutional responsibility for addressing gender-based violence is understood.

	Continuous Distance Measures				Binary Distance Measure	
	(1)	(2)	(3)	(4)	(5)	(6)
Info Social Norm (Treated)	0.055 (0.942)	0.081 (0.894)	0.097 (0.894)	0.068 (1.309)	0.161 (0.954)	1.589 (1.808)
Info Social Norm $\times$ Woman				0.059 (1.791)		
Woman		12.052*** (0.899)	12.299*** (0.904)	12.266*** (1.336)	12.309*** (0.905)	11.833*** (0.903)
Distance Social Norm_Study Abroad			-0.0477** (0.0215)	-0.0477** (0.0215)	-0.0657** (0.0320)	-3.183** (1.483)
Info Social Norm $\times$ Distance Social Norm_Study Abroad					0.028 (0.043)	0.705 (1.966)
Distance Social Norm_Private Picture			-0.0079 (0.0194)	-0.0079 (0.0194)	-0.0150 (0.028)	-0.951 (1.437)
Info Social Norm $\times$ Distance Social Norm_Private Picture					0.014 (0.038)	-0.205 (1.895)
Distance Social Norm_Share Location			0.0099 (0.0201)	0.0099 (0.0201)	0.0159 (0.030)	-2.885* (1.496)
Info Social Norm $\times$ Distance Social Norm_Share Location					-0.012 (0.040)	2.064 (1.947)
Distance Social Norm_Social Media			-0.0006 (0.0191)	-0.0006 (0.0191)	0.034 (0.030)	2.296 (1.481)
Info Social Norm $\times$ Distance Social Norm_Social Media					-0.059 (0.039)	-5.033** (1.942)
Distance_Study Abroad (abs. value)						0.0159 (0.032)
Distance_Private Picture (abs. value)						0.0665** (0.031)
Distance_Share Location (abs. value)						0.1025*** (0.032)
Distance_Social Media (abs. value)						0.0434 (0.031)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports OLS coefficients for the effect of the information treatment on respondents' support for *sexual education*. The dependent variable is measured on a 0–100 scale. Continuous and binary distance measures capture the deviation between perceived and actual social norms for each vignette. Column (1) shows bivariate results; (2) adds controls; (3)–(5) include continuous distance measures and interactions; (6) replaces them with binary indicators and absolute distances. Models (2) to (6) include a set of controls: country of residence, religiosity, student status, education level, mother's education and employment, relationship status, and being born abroad.

Table 11: Effect of Information Treatment on Support for Sexual Education

	Continuous Distance Measures					Binary Distance Measure
	(1)	(2)	(3)	(4)	(5)	(6)
Info Social Norm (Treated)	-1.286 (0.892)	-1.190 (0.873)	-1.166 (0.873)	-2.134* (1.252)	-0.926 (0.925)	0.787 (1.748)
Info Social Norm $\times$ Woman				1.936 (1.748)		
Woman		9.455*** (0.881)	9.764*** (0.887)	8.689*** (1.303)	9.788*** (0.888)	9.219*** (0.883)
Distance Social Norm_Study Abroad			-0.017 (0.021)	-0.018 (0.021)	-0.004 (0.030)	-1.196 (1.444)
Info Social Norm $\times$ Distance Social Norm_Study Abroad					-0.027 (0.041)	-1.711 (1.918)
Distance Social Norm_Private Picture			-0.035* (0.019)	-0.035* (0.019)	-0.049* (0.027)	-2.615* (1.378)
Info Social Norm $\times$ Distance Social Norm_Private Picture					0.028 (0.037)	0.220 (1.853)
Distance Social Norm_Share Location			-0.003 (0.020)	-0.003 (0.020)	0.021 (0.030)	-1.996 (1.456)
Info Social Norm $\times$ Distance Social Norm_Share Location					-0.044 (0.040)	1.084 (1.924)
Distance Social Norm_Social Media			-0.000 (0.019)	-0.000 (0.019)	0.006 (0.030)	1.165 (1.438)
Info Social Norm $\times$ Distance Social Norm_Social Media					-0.012 (0.039)	-2.685 (1.908)
Distance_Study Abroad (abs. value)						0.0368 (0.032)
Distance_Private Picture (abs. value)						0.1022*** (0.031)
Distance_Share Location (abs. value)						0.0755** (0.032)
Distance_Social Media (abs. value)						0.0421 (0.031)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports OLS estimates for the effect of the information treatment on respondents' support for *awareness campaigns against gender-based violence*, measured on a 0–100 scale. Continuous and binary distance measures capture the deviation between perceived and actual social norms for each vignette. Columns (1)–(5) include continuous measures and their interactions; column (6) includes binary and absolute distance measures. Models (2) to (6) include a set of controls: country of residence, religiosity, student status, education level, mother's education and employment, relationship status, and being born abroad.

Table 12: Effect of Information Treatment on Support for Awareness Campaigns

	Continuous Distance Measures					Binary Distance Measure
	(1)	(2)	(3)	(4)	(5)	(6)
Info Social Norm (Treated)	-0.113 (0.923)	-0.133 (0.908)	-0.083 (0.906)	1.747 (1.307)	-0.506 (0.954)	-2.568 (1.809)
Info Social Norm $\times$ Woman				-3.658** (1.811)		
Woman		8.902*** (0.913)	9.356*** (0.916)	11.387*** (1.338)	9.297*** (0.917)	8.602*** (0.910)
Distance Social Norm_Study Abroad			-0.040* (0.021)	-0.039* (0.021)	-0.047 (0.032)	-1.641 (1.467)
Info Social Norm $\times$ Distance Social Norm_Study Abroad					0.018 (0.043)	1.096 (1.982)
Distance Social Norm_Private Picture			-0.057*** (0.019)	-0.057*** (0.019)	-0.054* (0.028)	-3.207** (1.413)
Info Social Norm $\times$ Distance Social Norm_Private Picture					-0.008 (0.038)	-0.563 (1.904)
Distance Social Norm_Share Location			0.019 (0.020)	0.018 (0.020)	0.012 (0.029)	-4.626*** (1.474)
Info Social Norm $\times$ Distance Social Norm_Share Location					0.015 (0.040)	5.682*** (1.978)
Distance Social Norm_Social Media			0.018 (0.019)	0.018 (0.019)	-0.018 (0.029)	0.085 (1.481)
Info Social Norm $\times$ Distance Social Norm_Social Media					0.065* (0.039)	-1.257 (1.971)
Distance_Study Abroad (abs. value)						0.0298 (0.032)
Distance_Private Picture (abs. value)						0.0564* (0.032)
Distance_Share Location (abs. value)						0.1312*** (0.033)
Distance_Social Media (abs. value)						0.0958*** (0.031)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports OLS estimates for the effect of the information treatment on respondents' support for *harsher punishment for perpetrators of gender-based violence*. The dependent variable is measured on a 0–100 scale. Continuous and binary distance measures capture the deviation between perceived and actual social norms for each vignette. Columns (1)–(5) include continuous measures and their interactions; column (6) replaces them with binary indicators and absolute distances. Models (2) to (6) include a set of controls: country of residence, religiosity, student status, education level, mother's education and employment, relationship status, and being born abroad.

Table 13: Effect of Information Treatment on Support for Harsher Punishment for Perpetrators

	Continuous Distance Measures					Binary Distance Measure
	(1)	(2)	(3)	(4)	(5)	(6)
Info Social Norm (Treated)	-0.202 (0.909)	-0.091 (0.886)	-0.085 (0.885)	1.145 (1.245)	-0.388 (0.960)	-0.572 (1.852)
Info Social Norm $\times$ Woman				-2.458 (1.775)		
Woman		7.803*** (0.893)	7.660*** (0.897)	9.026*** (1.316)	7.649*** (0.898)	7.420*** (0.889)
Distance Social Norm_Study Abroad			-0.004 (0.021)	-0.003 (0.021)	-0.021 (0.032)	-2.559* (1.461)
Info Social Norm $\times$ Distance Social Norm_Study Abroad					0.027 (0.042)	2.154 (1.950)
Distance Social Norm_Private Picture			0.011 (0.019)	0.011 (0.019)	-0.014 (0.028)	-1.645 (1.390)
Info Social Norm $\times$ Distance Social Norm_Private Picture					0.048 (0.038)	1.956 (1.867)
Distance Social Norm_Share Location			-0.006 (0.020)	-0.006 (0.020)	-0.018 (0.030)	-3.252** (1.455)
Info Social Norm $\times$ Distance Social Norm_Share Location					0.021 (0.040)	3.581* (1.933)
Distance Social Norm_Social Media			0.054*** (0.019)	0.054*** (0.019)	0.093*** (0.028)	5.674*** (1.442)
Info Social Norm $\times$ Distance Social Norm_Social Media					-0.067* (0.038)	-6.318*** (1.909)
Distance_Study Abroad (abs. value)						0.0732** (0.032)
Distance_Private Picture (abs. value)						0.1076*** (0.031)
Distance_Share Location (abs. value)						0.0813** (0.032)
Distance_Social Media (abs. value)						0.1073*** (0.030)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports OLS estimates for the effect of the information treatment on respondents' *support for anti-violence centers and related initiatives*. The dependent variable is measured on a 0–100 scale. Continuous and binary distance measures capture the deviation between perceived and actual social norms for each vignette. Columns (1)–(5) include continuous measures and their interactions; column (6) replaces them with binary indicators and absolute distances. Models (2) to (6) include a set of controls: country of residence, religiosity, education, student status, mother's education and employment, relationship status, and being born abroad.

Table 14: Effect of Information Treatment on Support for Anti-Violence Centers

7 Conclusion

This study examined how correcting misperceptions of social norms about controlling behaviours affects young adults’ personal beliefs, behavioural intentions, and policy preferences related to gender-based violence. Across four vignettes depicting everyday forms of coercive control, we find that respondents systematically overestimate their peers’ tolerance toward such behaviours. Providing information about the actual social norm might help reduce this bias, but its effects are heterogeneous. In particular, men with larger positive misperceptions (that is, those who overestimate the share of their peers who would recommend the coercive option relative to the actual social norm) respond most strongly to corrective feedback, revising their personal recommendations away from coercive options, while women, who already hold more progressive baselines, show limited/ non significant updating.

Beyond attitudinal change, the treatment has nuanced effects on broader outcomes. Informational corrections do not uniformly increase support for institutional measures; rather, they reshape how respondents perceive the need for intervention. When individuals learn that peers disapprove of controlling behaviours, support for punitive measures tends to rise, while support for preventive or support-based initiatives sometimes declines, suggesting that perceived social prevalence influences perceived policy urgency.

Turning to the behavioural dimension, we tested whether these attitudinal shifts translated into actual behaviour by analysing participants’ donations to the NGO Women Against Violence Europe (WAVE). While nearly all respondents chose to donate, the information treatment did not significantly affect either the probability or the amount donated, suggesting limited behavioural spillover. Women were more likely to donate and contributed larger amounts on average, confirming gendered differences in prosocial engagement. For respondents who overestimated their peers’ tolerance toward coercive behaviour, however, there is a marginally significant treatment effect, indicating an increase in the probability of donating, and among full-time workers—those presumably less financially constrained—the interaction between treatment and misperception becomes significant for the donation amount. These patterns suggest that correcting normative misperceptions may enhance behavioural support for anti-violence initiatives, but that financial or motivational constraints can attenuate these

effects. In this regard, further analysis is needed, particularly through the examination of participants' open-ended responses explaining the motivations behind their donation choices.

The results underscore that informational interventions primarily shape values and perceived norms, while their translation into concrete behavioural change depends on contextual and individual factors. Among young Europeans, misperceptions about what others tolerate contribute to sustaining harmful norms even in the absence of overt violence. Making visible the true social disapproval of controlling behaviours can thus complement education and prevention policies—but these strategies must consider heterogeneity in prior beliefs, resource constraints, and the symbolic meaning of each domain of intervention. In this sense, engaging young adults is essential: they are not only the most exposed to emerging, technology-mediated forms of control, but also the most receptive agents of cultural change. Tackling gender-based violence therefore requires acting on both its visible manifestations and its invisible social roots.

Acknowledgments

This study was conducted with the internal funding of the Joint Research Centre (JRC) of the European Commission.

References

- Hunt Allcott. Social norms and energy conservation. *Journal of public Economics*, 95(9-10): 1082–1095, 2011.
- Mark A Andor, Andreas Gerster, Jörg Peters, and Christoph M Schmidt. Social norms and energy conservation beyond the us. *Journal of Environmental Economics and Management*, 103:102351, 2020.
- Peter Andre, Teodora Boneva, Felix Chopra, and Armin Falk. Misperceived social norms and willingness to act against climate change. *Review of Economics and Statistics*, pages 1–46, 2024.
- Ian Ayres, Sophie Raseman, and Alice Shih. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, & Organization*, 29(5):992–1022, 2013.
- Francesca Barigozzi and Natalia Montinari. Social norms in survey experiments: Personal beliefs versus normative expectations. *Journal of Economic Behavior & Organization*, 239: 107292, 2025.
- Sevinc Bermek, Konstantinos Matakos, and Asli Unan. Victim-blaming social norms and violence against women: Correcting misperceptions or morality drive policy and behavior change? *Available at SSRN 4609063*, 2023.
- Maria Bernedo, Paul J Ferraro, and Michael Price. The persistent impacts of norm-based messaging and their implications for water conservation. *Journal of Consumer Policy*, 37 (3):437–452, 2014.

- Cristina Bicchieri. *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press, 2016.
- Cristina Bicchieri and Erte Xiao. Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2):191–208, 2009.
- Leonardo Bursztyn and David Y Yang. Misperceptions about others. *Annual Review of Economics*, 14(1):425–452, 2022.
- Leonardo Bursztyn, Alessandra L González, and David Yanagizawa-Drott. Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, 110(10):2997–3029, 2020.
- Patricia Cortés, Gizem Koşar, Jessica Pan, and Basit Zafar. Should mothers work? how perceptions of the social norm affect individual attitudes toward work in the us. *Review of Economics and Statistics*, pages 1–28, 2024.
- Dora L Costa and Matthew E Kahn. Energy conservation “nudges” and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702, 2013.
- Council of Europe. Convention on preventing and combating violence against women and domestic violence. Council of Europe Treaty Series No. 210 (“Istanbul Convention”), 2011. URL <https://rm.coe.int/168008482e>.
- European Commission. A union of equality: Gender equality strategy 2020–2025. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2020. URL <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0152>.
- European Commission. What is gender-based violence? https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence/what-gender-based-violence_en, 2025. Accessed: 22 Oct 2025.

European Institute for Gender Equality. The costs of gender-based violence in the european union, 2021. URL <https://eige.europa.eu/publications/costs-gender-based-violence-european-union>. Report.

European Institute for Gender Equality. Gender equality index 2024: Sustaining momentum on a fragile path, 2024. URL <https://eige.europa.eu/publications-resources/publications/gender-equality-index-2024-sustaining-momentum-fragile-path>. Accessed: 23 Oct 2025.

European Institute for Gender Equality. Psychological violence. https://eige.europa.eu/publications-resources/thesaurus/terms/1241?language_content_entity=en, n.d.a. Accessed: 23 Oct 2025.

European Institute for Gender Equality. Coercive control. <https://eige.europa.eu/publications-resources/thesaurus/terms/1422>, n.d.b. Accessed: 23 Oct 2025.

European Parliament and Council. Directive (eu) 2024/1385 of 14 may 2024 on combating violence against women and domestic violence, 2024. URL <https://eur-lex.europa.eu/eli/dir/2024/1385/oj>. Official Journal of the European Union, 24 May 2024. CELEX: 32024L1385.

Paul J Ferraro and Michael K Price. Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1): 64–73, 2013.

Hanna Fuhrmann-Riebel, Ben D’Exelle, Kristian López Vargas, Sebastian Tonke, and Arjan Verschoor. Correcting misperceptions about trends and norms to address weak collective action—experimental evidence from a recycling program. *Journal of Environmental Economics and Management*, 128:103046, 2024.

Claudia Garcia-Moreno, Henrica AFM Jansen, Mary Ellsberg, Lori Heise, and Charlotte H Watts. Prevalence of intimate partner violence: findings from the who multi-country study on women’s health and domestic violence. *The lancet*, 368(9543):1260–1269, 2006.

- Noah J Goldstein, Robert B Cialdini, and Vlasdas Griskevicius. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, 35(3):472–482, 2008.
- Michael Hallsworth, John A List, Robert D Metcalfe, and Ivo Vlaev. The behaviorist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of public economics*, 148:14–31, 2017.
- N Zoe Hilton and Dana L Radatz. Coercive control in a national us self-report survey: Prediction of repeated intimate partner violence. *Psychological services*, 2024.
- Erin L Krupka and Roberto A Weber. Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524, 2013.
- Ieda Matavelli. We don’t talk about boys: Masculinity norms among adolescents in brazil. *University of British Columbia, mimeo*, 2024.
- Joseph S Mernyk, Sophia L Pink, James N Druckman, and Robb Willer. Correcting inaccurate metaperceptions reduces americans’ support for partisan violence. *Proceedings of the National Academy of Sciences*, 119(16):e2116851119, 2022.
- Alex Moehring, Avinash Collis, Kiran Garimella, M Amin Rahimian, Sinan Aral, and Dean Eckles. Providing normative information increases intentions to accept a covid-19 vaccine. *Nature Communications*, 14(1):126, 2023.
- Clayton Neighbors, Mary E Larimer, and Melissa A Lewis. Targeting misperceptions of descriptive drinking norms: efficacy of a computer-delivered personalized normative feedback intervention. *Journal of consulting and clinical psychology*, 72(3):434, 2004.
- P Wesley Schultz, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, and Vlasdas Griskevicius. The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5):429–434, 2007.

Jen Shang and Rachel Croson. A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The economic journal*, 119(540):1422–1439, 2009.

Margaret E Tankard and Elizabeth Levy Paluck. Norm perception as a vehicle for social change. *Social issues and policy review*, 10(1):181–211, 2016.

Jan G Voelkel, Michael N Stagnaro, James Y Chu, Sophia L Pink, Joseph S Mernyk, Chrystal Redekopp, Isaias Ghezae, Matthew Cashman, Dhaval Adjodah, Levi G Allen, et al. Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity. *Science*, 386(6719):eadh4764, 2024.

Appendix A

Vignettes (English version)

In the design phase, we developed six vignettes covering different domains of partner control and social norms (Q21–Q32)¹⁰. To avoid making the survey excessively burdensome for respondents, in the pilot stage each participant was randomly assigned four vignettes out of the full set of six. After analysing the pilot data, we decided to retain only four vignettes for the main study. The selection was based on two criteria.

First, we focused on the vignettes that generated the largest differences between second-order beliefs (SOB) and first-order beliefs (FOB), as these provided the clearest evidence of potential misperceptions of social norms. Second, we considered the distribution of responses and excluded vignettes that displayed the highest share of extreme values at 0 or 100, which suggested limited variation and potential ceiling or floor effects.

The final set of vignettes included in the main survey are: Control over Opportunity (Q21), Sharing without Consent (Q23), Control over Behaviour (Q27), and Social Media Control (Q29).

Q21 – Control over Opportunity

Martha, a university student, has been accepted for an Erasmus programme that will allow her to study abroad for one semester. She is excited about the idea of experiencing something new, having fun and meeting new people. However, this will mean being away from home for several months. Martha has been in a stable relationship with Luke for more than a year and tells him the news. Luke is concerned about the distance and how this experience might affect their relationship.

We asked other women/men from your age group and country how they think someone should react in Luke’s situation. In your opinion, what percentage of them would advise Luke to behave in each of the following ways?

¹⁰The vignettes were translated into Swedish and Romanian using a translation and back-translation procedure, and the characters’ names were adapted to common names in each country.

Option A: Luke tells Martha that he trusts her, but does not feel comfortable with the idea of her being with new people in a new environment. He asks her to think it over, telling her that he would miss her greatly and would prefer that she did not leave.

Option B: Luke acknowledges that being apart will be difficult, but fully supports Martha's decision. He suggests organising regular calls and visits, seeing this period apart as an opportunity to grow.

Q22 – Control over Opportunity (FOB)

Before continuing, we'd like to tell you how participants who completed a previous version of this questionnaire responded.

Among the men/women from your age group and country, [Y]% would advise Luke to ask Martha to reconsider and not leave (Option A).

What advice would you give to Luke? Specifically, what is the probability (out of 100) of your recommending each of the following options?

(Note: the percentages do not have to add up to 100)

Q23 – Sharing without Consent

One evening out with friends, David is scrolling through the photos on his phone. Laughing, he shows the others a picture of his ex-womanfriend in a swimsuit, taken on holiday when they were still together. As he shows it, he jokingly remarks: "Those were the days... get a load of that!" Some friends laugh, others make jokes. Alexander, who is sitting nearby, observes the scene. It bothers him, but he is not sure whether to speak out or let it be.

We asked other women/men from your age group and country how they think someone should react in Alexander's situation. In your opinion, what percentage of them would advise Alexander to behave in each of the following ways?

Option A: Alexander smiles slightly, but he feels annoyed. He doesn't approve of what David did, but dismisses it as a joke in a light-hearted situation. He decides not to say anything as no one else seems upset by the situation.

Option B: Without raising his voice, Alexander says something like: “Come on, you shouldn’t really be showing us those photos... it’s a private matter.” He doesn’t confront anyone, but makes it clear that the behaviour made him uncomfortable.

Q24 – Sharing without Consent (FOB)

Before continuing, we’d like to tell you how participants who completed a previous version of this questionnaire responded.

Among the men/women from your age group and country, [Y]% would advise Alexander to mention that the behaviour made him uncomfortable (Option B).

What advice would you give to Alexander? Specifically, what is the probability (out of 100) of your recommending each of the following options?

(Note: the percentages do not have to add up to 100)

Q25 – Social Media Control (Lucy)

Lucy is 20 years old and passionate about photography and social media. She still follows one of her ex-boyfriends on Instagram. From time to time, she reacts to his stories with an emoji or brief comment. One afternoon, while searching for a meme on Lucy’s phone, her current partner William notices that her ex is often listed among her recent contacts and that they’ve had several interactions. This makes him feel uncomfortable.

We asked other women/men from your age group and country how they think someone should react in William’s situation. In your opinion, what percentage of them would advise William to behave in each of the following ways?

Option A: William calmly asks Lucy to block her ex on Instagram, explaining that it would be a simple way to help him feel more at ease.

Option B: William suggests to Lucy that they agree for a time that each can freely check the other’s phone whenever they want, for mutual reassurance.

Q26 – Social Media Control (FOB)

Before continuing, we'd like to tell you how participants who completed a previous version of this questionnaire responded.

Among the men/women from your age group and country, [Y]% would advise William to suggest agreeing that they can look at each other's phones (Option B).

What advice would you give to William? Specifically, what is the probability (out of 100) of your recommending each of the following options?

(Note: the percentages do not have to add up to 100)

Q27 – Control over Behaviour

Carl is 24 years old and lives separately from his womanfriend Martha during the week. Martha goes to university in a different city from where Carl lives, while he works in the city where they both grew up. Sometimes Carl feels insecure and wonders what Martha is doing when they are not in touch.

We asked other men/women from your age group and country how they think someone should react in Carl's situation. In your opinion, what percentage of them would advise Carl to behave in each of the following ways?

Option A: Carl suggests that Martha should let him know her location so they can feel closer while they're apart.

Option B: Carl suggests arranging a daily evening phone call, so they can speak to each other regularly.

Q28 – Control over Behaviour (FOB)

Before continuing, we'd like to tell you how participants who completed a previous version of this questionnaire responded.

Among the men/women from your age group and country, [Y]% would advise Carl to suggest that Martha let him know her location (Option A).

What advice would you give to Carl? Specifically, what is the probability (out of 100) of your recommending one of the following options?

(Note: the percentages do not have to add up to 100)

Q29 – Social Media Control (Julia)

Julia is 22 years old and has just got back from a holiday in Greece. She posted some photos of herself wearing a swimsuit on Instagram. While scrolling through the comments, her boyfriend Mark noticed several suggestive emojis and innuendos, which made him feel uncomfortable.

We asked other men/women from your age group and country how they think someone should react in Mark's situation. In your opinion, what percentage of them would advise Mark to behave in each of the following ways?

Option A: Mark asks Julia to delete the photos of herself in a swimsuit, explaining that the comments make him uncomfortable.

Option B: Mark accepts that the photos remain online but asks her to avoid posting similar ones in future due to the inappropriate comments they attract.

Q30 – Social Media Control (FOB)

Before continuing, we'd like to tell you how participants who completed a previous version of this questionnaire responded.

Among the men/women from your age group and country, [Y]% would advise Mark to ask Julia to delete the photos (Option A).

What advice would you give to Mark? Specifically, what is the probability (out of 100) of your recommending each of the following options?

(Note: the percentages do not have to add up to 100)

Q31 – Going Out (Emma)

Emma, 23 years old, is a fan of electronic music and a new group of friends has invited her to a club night where a DJ she really likes is playing. She is keen on going, but her boyfriend Matt – with whom she has been in a steady relationship for 18 months – doesn't know these friends and feels a little uncomfortable with the idea of Emma being in an environment where people often meet and flirt.

We asked other men/women from your age group and country how they think someone should react in Matt's situation. In your opinion, what percentage of them would advise Matt to behave in each of the following ways?

Option A: Matt asks Emma not to go on the night out, explaining that nightclubs can be places where you get unwanted attention and that he would feel more comfortable if she stayed at home.

Option B: Matt says he has nothing against it, but as he doesn't know the group, he asks her if he can come along to get to know her new friends.

Q32 – Going Out (FOB)

Before continuing, we'd like to tell you how participants who completed a previous version of this questionnaire responded.

Among the men/women from your age group and country, [Y]% would advise Matt to ask Emma not to go on the night out (Option A).

What advice would you give to Matt? Specifically, what is the probability (out of 100) of your recommending each of the following options?

(Note: the percentages do not have to add up to 100)

Appendix B

Table B1: Distribution of Categorical Variables by Treatment Group (Counts)

Variable	Category	Control (N)	Treated (N)	Total (N)	p-value
Country	Italy	520	650	1,170	1.000
	Romania	520	650	1,170	
	Sweden	520	650	1,170	
Gender	Male	780	975	1,755	1.000
	Female	780	975	1,755	
Area (arean)	Metropolitan area	481	575	1,056	0.916
	Medium city	495	623	1,118	
	Small city	397	508	905	
	Rural area	184	241	425	
	Other/unspecified	3	3	6	
Born abroad (respondent)	No	1,470	1,838	3,308	0.974
	Yes	90	112	202	
Education (self)	Primary education	27	32	59	0.913
	Lower secondary	96	122	218	
	Upper secondary	574	695	1,269	
	Vocational / Professional	223	284	507	
	Some university (no degree)	239	279	518	
	University degree	305	403	708	
	Postgraduate (Master/PhD)	61	89	150	
	Prefer not to say / Other	32	45	77	
Religious	No	798	982	1,780	0.640
	Yes	762	968	1,730	
Religious denomination	Christian	667	847	1,514	0.580
	Muslim	77	89	166	
	Other	7	10	17	
	Prefer not to say (PNS)	11	22	33	
Employment	Full-time	524	707	1,231	0.115
	Part-time	224	233	457	
	Unemployed	180	194	374	
	Homemaker/Parent	44	51	95	
	Student	562	737	1,299	
	Other	26	28	54	
Relationship status	Single / not in a relationship	588	712	1,300	0.099
	In relationship, not cohabiting	516	636	1,152	
	In relationship, cohabiting	407	561	968	
	Prefer not to say	49	41	90	
Mother born abroad	No	1,341	1,680	3,021	0.870
	Yes	219	270	489	
Father born abroad	No	1,370	1,692	3,062	0.354
	Yes	190	258	448	
Father education	Primary or less	119	133	252	0.305
	Lower secondary	189	257	446	
	Vocational / upper secondary	716	856	1,572	
	Some university (no degree)	79	114	193	
	University or higher	363	491	854	
	Prefer not to say	94	99	193	
Mother education	Primary or less	109	118	227	0.303
	Lower secondary	93	114	207	
	Vocational / upper secondary	713	949	1,662	
	Some university (no degree)	118	117	235	
	University or higher	448	557	1,005	
	Prefer not to say	79	95	174	

Note: p-values are from Pearson chi-square tests.

Variable	Category	Control (N)	Treated (N)	Total (N)	p-value
Mother employment	Full-time	979	1,248	2,227	0.631
	Part-time	136	186	322	
	Unemployed	72	73	145	
	Homemaker/Parent	175	217	392	
	Student	79	94	173	
	Other	119	132	251	
Father employment	Full-time	1,005	1,316	2,321	0.037
	Part-time	111	154	265	
	Unemployed	81	63	144	
	Homemaker/Parent	43	48	91	
	Student	168	202	370	
	Other	152	167	319	

Note: p-values are from Pearson chi-square tests.

Characteristic	N	Pearson χ^2 (df)	p-value
Country	673	8.97 (2)	0.011
Gender	649	0.32 (1)	0.569
Area	584	2.53 (3)	0.469
Education	546	3.95 (7)	0.786
Religious	673	0.12 (1)	0.724
Employment	556	2.94 (5)	0.709
Relationship status	382	2.24 (3)	0.525
Father's education	405	0.18 (4)	0.996
Mother's education	403	3.60 (4)	0.463
Mother's employment	488	6.30 (5)	0.278
Father's employment	525	7.07 (5)	0.216

Table B2: Selective attrition checks across waves (attriters only)

Notes: The table reports Pearson chi-squared tests of independence between survey wave and each baseline characteristic, computed on the *attriting sample* (i.e., respondents who drop out).

Table B3: FOB, SOB, and misperception (SOB–Truth), control group

Vignette	Option	FOB mean (sd)	SOB mean (sd)	(SOB–Truth) mean	N	p (SOB–Truth)
Study Abroad	A	37.18 (28.99)	42.98 (26.96)	+5.80	1560	< 0.001
	B	72.93 (25.47)	67.78 (25.07)	–5.15	1560	< 0.001
Private Picture	A	38.82 (29.05)	45.67 (27.92)	+6.85	1560	< 0.001
	B	68.04 (28.13)	62.53 (28.34)	–5.51	1560	< 0.001
Share Location	A	37.42 (29.54)	42.52 (26.96)	+5.10	1560	< 0.001
	B	68.90 (26.14)	67.11 (23.98)	–1.80	1560	0.003
Social Media	A	42.07 (31.03)	46.68 (28.08)	+4.60	1560	< 0.001
	B	55.42 (29.85)	54.83 (27.31)	–0.59	1560	0.393

Notes: FOB = first-order belief (“what you would advise”); SOB = second-order belief (“share of your peers who would advise”). *Truth* is defined as the mean FOB within country $\times$ gender in the control group. p (SOB–Truth) reports one-sample t -tests of H_0 : mean(SOB–Truth) = 0 (two-sided).

Table B4: FOB, SOB, and misperception (SOB–Truth), by country, control group

Vignette	Option	FOB mean (sd)	SOB mean (sd)	(SOB–Truth) mean (sd)	N	<i>p</i> (SOB–Truth)
Italy						
Study Abroad	A	31.36 (28.92)	40.43 (27.20)	+9.07 (25.91)	520	< 0.001
	B	76.60 (25.00)	68.19 (24.74)	–8.40 (24.14)	520	< 0.001
Private Picture	A	38.86 (30.46)	46.21 (28.45)	+7.34 (26.99)	520	< 0.001
	B	68.20 (29.17)	61.95 (28.89)	–6.25 (28.33)	520	< 0.001
Share Location	A	31.37 (29.59)	38.36 (26.96)	+6.98 (26.57)	520	< 0.001
	B	69.20 (26.25)	67.00 (23.06)	–2.21 (22.95)	520	0.029
Social Media	A	35.85 (30.63)	44.01 (28.70)	+8.15 (27.68)	520	< 0.001
	B	50.51 (30.79)	51.13 (27.05)	+0.61 (27.24)	520	0.609
Romania						
Study Abroad	A	38.18 (28.96)	42.35 (27.73)	+4.17 (27.10)	520	< 0.001
	B	74.52 (25.57)	70.84 (26.11)	–3.68 (25.41)	520	0.001
Private Picture	A	37.67 (28.92)	42.00 (28.21)	+4.33 (27.64)	520	< 0.001
	B	70.80 (27.80)	68.09 (27.95)	–2.71 (27.65)	520	0.026
Share Location	A	37.24 (29.37)	42.66 (28.31)	+5.42 (28.08)	520	< 0.001
	B	73.12 (25.96)	71.36 (24.51)	–1.76 (24.47)	520	0.102
Social Media	A	44.73 (31.53)	46.65 (28.47)	+1.91 (28.15)	520	0.122
	B	61.44 (29.60)	59.90 (28.64)	–1.54 (28.57)	520	0.218
Sweden						
Study Abroad	A	41.99 (28.12)	46.16 (25.65)	+4.17 (25.14)	520	< 0.001
	B	67.68 (25.02)	64.32 (23.92)	–3.37 (23.47)	520	0.001
Private Picture	A	39.92 (27.71)	48.81 (26.71)	+8.88 (26.32)	520	< 0.001
	B	65.12 (27.15)	57.55 (27.22)	–7.57 (27.15)	520	< 0.001
Share Location	A	43.66 (28.40)	46.54 (24.93)	+2.88 (25.00)	520	0.009
	B	64.39 (25.50)	62.96 (23.66)	–1.43 (23.54)	520	0.166
Social Media	A	45.63 (30.02)	49.38 (26.83)	+3.75 (26.26)	520	0.001
	B	54.31 (28.11)	53.48 (25.45)	–0.83 (25.44)	520	0.458

Notes: FOB = first-order belief (“what you would advise”); SOB = second-order belief (“share of your peers who would advise”). *Truth* is defined as the mean FOB within country $\times$ gender in the control group. *p* values from one-sample *t* tests of H_0 : mean(SOB–Truth) = 0 (two-sided).

Table B5: Gender- and country-specific FOB, SOB, (SOB–Truth), and significance by vignette option, control group

Vignette	Option	Italy				Romania				Sweden			
		FOB	SOB	SOB–Truth	<i>p</i>	FOB	SOB	SOB–Truth	<i>p</i>	FOB	SOB	SOB–Truth	<i>p</i>
Women													
Study Abroad	A	21.22	31.97	+10.75	< 0.001	31.41	36.46	+5.05	0.002	34.32	40.62	+6.30	< 0.001
	B	83.34	73.72	−9.62	< 0.001	80.33	76.80	−3.53	0.016	72.99	68.98	−4.01	0.009
Private Picture	A	24.66	36.26	+11.60	< 0.001	31.00	36.29	+5.29	0.002	31.72	43.46	+11.73	< 0.001
	B	79.22	68.89	−10.33	< 0.001	75.90	72.27	−3.63	0.034	71.79	61.18	−10.61	< 0.001
Share Location	A	24.75	33.46	+8.71	< 0.001	33.87	39.04	+5.17	0.005	39.88	45.10	+5.22	0.001
	B	71.38	69.19	−2.19	0.151	75.60	72.99	−2.62	0.089	68.44	65.67	−2.78	0.071
Social Media	A	27.62	36.42	+8.80	< 0.001	41.13	42.32	+1.19	0.498	37.83	43.55	+5.72	< 0.001
	B	45.16	49.41	+4.25	0.017	63.50	61.89	−1.60	0.372	55.96	54.38	−1.58	0.342
Men													
Study Abroad	A	41.50	48.88	+7.38	< 0.001	44.95	48.25	+3.30	0.055	49.66	51.70	+2.04	0.168
	B	69.85	62.67	−7.18	< 0.001	68.70	64.87	−3.83	0.025	62.38	59.66	−2.72	0.051
Private Picture	A	53.07	56.15	+3.09	0.061	44.34	47.70	+3.36	0.050	48.12	54.15	+6.03	< 0.001
	B	57.18	55.02	−2.17	0.188	65.69	63.91	−1.78	0.302	58.45	53.92	−4.53	0.004
Share Location	A	38.00	43.25	+5.26	0.001	40.60	46.27	+5.67	< 0.001	47.44	47.98	+0.54	0.721
	B	67.02	64.80	−2.22	0.095	70.63	69.73	−0.90	0.548	60.35	60.26	−0.08	0.951
Social Media	A	44.08	51.59	+7.51	< 0.001	48.34	50.97	+2.63	0.131	53.44	55.22	+1.78	0.248
	B	55.87	52.84	−3.03	0.056	59.39	57.91	−1.48	0.398	52.65	52.57	−0.08	0.959

Notes: Truth is defined as the mean FOB within country $\times$ gender in the control group; SOB is respondents' second-order belief. (SOB–Truth) is the difference between SOB and Truth. p values are from one-sample t tests of H_0 : mean(SOB–Truth) = 0 (two-sided).

Table B6: Gender differences in SOB, FOB, and (SOB–Truth), control group

Scenario	Option	Δ SOB (M–W)			Δ FOB (M–W)			Δ (SOB–Truth) (M–W)		
		Diff	SE	p	Diff	SE	p	Diff	SE	p
Study Abroad	A	+13.26	1.32	< 0.001	+16.39	1.41	< 0.001	−3.13	1.32	0.018
	B	−10.77	1.24	< 0.001	−11.91	1.25	< 0.001	+1.14	1.24	0.357
Private Picture	A	+14.00	1.37	< 0.001	+19.38	1.39	< 0.001	−5.38	1.36	< 0.001
	B	−9.83	1.41	< 0.001	−15.19	1.37	< 0.001	+5.36	1.40	< 0.001
Share Location	A	+6.63	1.36	< 0.001	+9.18	1.48	< 0.001	−2.54	1.35	0.059
	B	−4.35	1.21	< 0.001	−5.81	1.32	< 0.001	+1.46	1.20	0.223
Social Media	A	+11.83	1.39	< 0.001	+13.09	1.54	< 0.001	−1.26	1.39	0.364
	B	−0.79	1.38	0.569	+1.10	1.51	0.469	−1.88	1.37	0.170

Notes: Δ SOB = (SOB men − SOB women); Δ FOB = (FOB men − FOB women); Δ (SOB–Truth) = ((SOB–Truth) men − (SOB–Truth) women). SE and p -values come from two-sample t -tests with equal variances (two-sided).

Table B7: Country-pair differences in SOB, FOB, and (SOB–Truth), control group

		IT-RO			IT-SV			RO-SV		
Scenario	Option	Diff	SE	<i>p</i>	Diff	SE	<i>p</i>	Diff	SE	<i>p</i>
ΔSOB										
Study Abroad	A	-1.93	1.70	0.259	-5.73	1.64	< 0.001	-3.81	1.66	0.022
	B	-2.64	1.58	0.094	+3.88	1.51	0.010	+6.52	1.55	< 0.001
Private Picture	A	+4.21	1.76	0.017	-2.60	1.71	0.129	-6.81	1.70	< 0.001
	B	-6.13	1.76	< 0.001	+4.41	1.74	0.012	+10.54	1.71	< 0.001
Share Location	A	-4.30	1.71	0.012	-8.19	1.61	< 0.001	-3.88	1.65	0.019
	B	-4.36	1.48	0.003	+4.03	1.45	0.005	+8.39	1.49	< 0.001
Social Media	A	-2.64	1.77	0.136	-5.38	1.72	0.002	-2.73	1.72	0.112
	B	-8.77	1.73	< 0.001	-2.35	1.63	0.149	+6.42	1.68	< 0.001
ΔFOB										
Study Abroad	A	-6.82	1.79	< 0.001	-10.63	1.77	< 0.001	-3.81	1.77	0.032
	B	+2.08	1.57	0.185	+8.91	1.55	< 0.001	+6.83	1.57	< 0.001
Private Picture	A	+1.19	1.84	0.518	-1.06	1.81	0.558	-2.25	1.76	0.200
	B	-2.59	1.77	0.142	+3.08	1.75	0.078	+5.68	1.70	< 0.001
Share Location	A	-5.87	1.83	0.001	-12.29	1.80	< 0.001	-6.42	1.79	< 0.001
	B	-3.92	1.62	0.016	+4.81	1.61	0.003	+8.72	1.60	< 0.001
Social Media	A	-8.88	1.93	< 0.001	-9.78	1.88	< 0.001	-0.90	1.91	0.637
	B	-10.93	1.87	< 0.001	-3.79	1.83	0.038	+7.14	1.79	< 0.001
Δ(SOB-Truth)										
Study Abroad	A	+4.89	1.64	0.003	+4.90	1.58	0.002	+0.00	1.62	0.999
	B	-4.72	1.54	0.002	-5.04	1.48	< 0.001	-0.32	1.52	0.835
Private Picture	A	+3.02	1.69	0.075	-1.54	1.65	0.351	-4.56	1.67	0.007
	B	-3.54	1.74	0.042	+1.32	1.72	0.443	+4.86	1.70	0.004
Share Location	A	+1.56	1.70	0.357	+4.10	1.60	0.010	+2.54	1.65	0.124
	B	-0.45	1.47	0.762	-0.78	1.44	0.591	-0.33	1.49	0.825
Social Media	A	+6.24	1.73	< 0.001	+4.41	1.67	0.009	-1.83	1.69	0.278
	B	+2.16	1.73	0.213	+1.44	1.63	0.378	-0.72	1.68	0.670

Notes: Each block reports country-pair differences (Italy–Romania, Italy–Sweden, Romania–Sweden). Diff = mean(first country) − mean(second country). Standard errors and p -values come from two-sample t -tests with equal variances (two-sided), computed in the control group.

Table B8: Means of perceived social norms (SOB) by treatment group

	Control (Wave 1)	Treated (Wave 2)	<i>p</i> -value
Study Abroad			
Option A	42.98	43.20	0.806
Option B	67.78	66.11	0.047
Private Picture			
Option A	45.67	46.28	0.505
Option B	62.53	60.36	0.022
Share Location			
Option A	42.52	42.41	0.903
Option B	67.11	66.65	0.576
Social Media			
Option A	46.68	46.42	0.789
Option B	54.83	55.04	0.822
Observations	1,560	1,950	3,510

Notes: The table reports mean values of perceived social norms (SOB) by treatment group (treated = 1 corresponds to Wave 2). *p*-values are from two-sample *t*-tests with equal variances (two-sided).

Appendix C

Table C1: FOB, SOB, and misperception (SOB–Truth), treated group

Vignette	Option	FOB mean (sd)	SOB mean (sd)	(SOB–Truth) mean	N	p (SOB–Truth)
Study Abroad	A	36.71 (26.84)	43.20 (25.50)	+6.02	1950	< 0.001
	B	70.90 (25.67)	66.11 (24.61)	–6.83	1950	< 0.001
Private Picture	A	40.46 (27.19)	46.28 (26.18)	+7.46	1950	< 0.001
	B	64.56 (27.64)	60.36 (27.70)	–7.68	1950	< 0.001
Share Location	A	38.35 (27.46)	42.41 (26.16)	+4.99	1950	< 0.001
	B	68.50 (25.42)	66.65 (24.24)	–2.26	1950	< 0.001
Social Media	A	41.99 (29.24)	46.42 (27.82)	+4.35	1950	< 0.001
	B	55.08 (28.84)	55.04 (27.06)	–0.38	1950	0.533

Notes: FOB = first-order belief (“what you would advise”); SOB = second-order belief (“share of your peers who would advise”). *Truth* is defined as the mean FOB within country $\times$ gender in the control group. p (SOB–Truth) reports one-sample t -tests of H_0 : mean(SOB–Truth) = 0 (two-sided), computed in the treated group.

	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se	(6) b/se
Info Social Norm	-0.467 (0.953)	-0.433 (0.905)	-0.564 (0.697)	-2.522** (1.025)	-0.225 (0.651)	3.831*** (1.012)
Distance Social Norm			0.660*** (0.015)	0.659*** (0.015)	0.691*** (0.022)	
Woman		-12.979*** (0.898)	-16.089*** (0.697)	-18.265*** (1.036)	-16.038*** (0.697)	-12.875*** (0.753)
Info Social Norm × Woman				3.915** (1.390)		
Info Social Norm × Distance					-0.057* (0.031)	
Overestimate						31.182*** (1.097)
Info Social Norm × Overestimate						-6.218*** (1.476)
Distance Social Norm (abs)						0.175*** (0.030)
Romania		6.008*** (1.206)	10.243*** (0.901)	10.228*** (0.900)	10.230*** (0.900)	8.133*** (0.986)
Sweden		7.757*** (1.151)	12.359*** (0.904)	12.329*** (0.904)	12.324*** (0.903)	9.265*** (0.972)
Religious		5.665*** (0.971)	3.156*** (0.748)	3.131*** (0.747)	3.191*** (0.748)	4.213*** (0.806)
University		-5.445*** (1.003)	-3.980*** (0.782)	-4.005*** (0.780)	-3.990*** (0.782)	-4.193*** (0.849)
Student		-2.954*** (0.955)	-1.924*** (0.730)	-1.911*** (0.729)	-1.951*** (0.730)	-2.493*** (0.787)
Born abroad		1.957 (1.964)	0.265 (1.558)	0.333 (1.553)	0.246 (1.559)	0.574 (1.692)
Metropolitan area		-0.487 (1.148)	0.174 (0.890)	0.170 (0.889)	0.216 (0.890)	0.226 (0.961)
Other/unspecified area		-2.821 (14.028)	0.896 (4.350)	0.882 (4.438)	1.447 (4.540)	-1.151 (8.809)
Rural area		2.303 (1.597)	0.522 (1.210)	0.496 (1.207)	0.589 (1.206)	1.423 (1.302)
Small city		-0.623 (1.169)	-0.389 (0.915)	-0.472 (0.913)	-0.335 (0.916)	0.213 (0.985)
Mother employed		-1.782* (1.069)	-1.499* (0.823)	-1.495* (0.822)	-1.543* (0.824)	-1.316 (0.883)
With partner		2.385** (0.938)	1.020 (0.727)	1.073 (0.726)	0.979 (0.726)	1.267 (0.778)
Mother university		1.149 (1.001)	0.308 (0.782)	0.338 (0.781)	0.303 (0.781)	0.360 (0.842)
Constant	37.176*** (0.734)	38.885*** (1.658)	34.887*** (1.250)	35.990*** (1.318)	34.718*** (1.243)	16.347*** (1.568)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: Determinants of Erasmus personal beliefs (OLS)

	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se	(6) b/se
Info Social Norm	1.639*	1.739*	1.385*	-0.288	1.381**	3.029***
Distance Social Norm	(0.959)	(0.913)	(0.747)	(1.083)	(0.695)	(1.099)
Info Social Norm $\times$ Woman			0.588***	0.587***	0.587***	
			(0.016)	(0.016)	(0.024)	
Info Social Norm $\times$ Distance				3.345**		
				(1.493)		
Positive distance dummy					0.000	26.979***
					(0.033)	(1.186)
Info Social Norm $\times$ Positive distance						-2.722*
						(1.571)
Absolute distance						0.122***
						(0.031)
Woman		-16.642***	-20.309***	-22.169***	-20.309***	-17.004***
		(0.913)	(0.739)	(1.135)	(0.739)	(0.793)
Romania		-0.245	1.233	1.225	1.233	1.083
		(1.223)	(0.968)	(0.966)	(0.968)	(1.036)
Sweden		1.241	0.735	0.716	0.735	0.699
		(1.148)	(0.967)	(0.966)	(0.967)	(1.012)
Religious		2.854***	3.106***	3.080***	3.106***	3.149***
		(0.984)	(0.804)	(0.803)	(0.804)	(0.850)
University		-1.734*	-2.102**	-2.121**	-2.102**	-1.816**
		(0.997)	(0.824)	(0.823)	(0.824)	(0.869)
Student		-1.200	-2.244**	-2.230**	-2.244**	-2.178**
		(0.969)	(0.795)	(0.795)	(0.795)	(0.843)
Born abroad		-0.188	-1.089	-1.031	-1.088	-0.329
		(1.998)	(1.726)	(1.723)	(1.727)	(1.770)
Metropolitan area		1.249	0.058	0.057	0.058	-0.022
		(1.151)	(0.945)	(0.943)	(0.945)	(1.001)
Other/unspecified area		18.372	14.703	14.700	14.704	11.408
		(14.099)	(12.133)	(12.301)	(12.135)	(12.922)
Rural area		3.000*	2.409*	2.385*	2.409*	2.801**
		(1.565)	(1.250)	(1.247)	(1.251)	(1.343)
Small city		2.443**	2.314**	2.244**	2.314**	2.326**
		(1.193)	(0.979)	(0.979)	(0.980)	(1.043)
Mother employed		-2.555**	-2.091**	-2.088**	-2.091**	-1.997**
		(1.082)	(0.871)	(0.871)	(0.870)	(0.935)
With partner		0.478	-0.388	-0.343	-0.388	-0.235
		(0.940)	(0.774)	(0.775)	(0.775)	(0.825)
Mother university		2.891***	2.699***	2.724***	2.699***	2.168**
		(1.004)	(0.815)	(0.815)	(0.815)	(0.867)
Constant	38.819***	45.646***	44.299***	45.237***	44.300***	28.404***
	(0.735)	(1.690)	(1.343)	(1.405)	(1.339)	(1.647)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: Determinants of Revenge Porn personal beliefs (OLS)

	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se	(6) b/se
Info Social Norm	0.930 (0.973)	0.901 (0.931)	1.030 (0.699)	-0.444 (0.994)	1.411** (0.657)	4.537*** (1.032)
Distance Social Norm			0.691*** (0.015)	0.691*** (0.015)	0.732*** (0.022)	
Info Social Norm $\times$ Woman				2.945** (1.400)		
Info Social Norm $\times$ Distance					-0.075** (0.030)	
Positive distance dummy						32.546*** (1.124)
Info Social Norm $\times$ Positive distance						-5.700*** (1.510)
Absolute distance						0.224*** (0.031)
Woman		-7.086*** (0.934)	-9.033*** (0.698)	-10.674*** (1.046)	-9.023*** (0.697)	-7.307*** (0.763)
Romania		4.931*** (1.236)	6.772*** (0.903)	6.766*** (0.902)	6.761*** (0.903)	5.960*** (0.984)
Sweden		11.191*** (1.198)	14.313*** (0.926)	14.295*** (0.925)	14.359*** (0.924)	12.116*** (1.002)
Religious		7.466*** (0.999)	4.719*** (0.765)	4.697*** (0.765)	4.731*** (0.765)	5.270*** (0.816)
University		-6.156*** (1.037)	-2.894*** (0.792)	-2.911*** (0.790)	-2.832*** (0.792)	-3.265*** (0.852)
Student		-3.092*** (0.984)	-2.333*** (0.735)	-2.322*** (0.734)	-2.368*** (0.735)	-2.079** (0.803)
Born abroad		3.438* (2.036)	0.195 (1.571)	0.245 (1.562)	0.207 (1.571)	1.369 (1.704)
Metropolitan area		-0.295 (1.188)	0.604 (0.901)	0.602 (0.901)	0.564 (0.899)	0.328 (0.972)
Other/unspecified area		-13.043 (8.502)	-1.735 (3.925)	-1.743 (4.062)	-0.969 (3.975)	-4.647 (6.307)
Rural area		0.013 (1.607)	1.136 (1.221)	1.114 (1.220)	1.130 (1.224)	1.107 (1.343)
Small city		-0.539 (1.227)	-0.230 (0.908)	-0.292 (0.907)	-0.202 (0.907)	0.016 (0.993)
Mother employed		0.210 (1.088)	-0.410 (0.815)	-0.406 (0.814)	-0.460 (0.814)	-0.377 (0.889)
With partner		2.866*** (0.977)	0.923 (0.737)	0.961 (0.737)	0.873 (0.736)	1.525* (0.803)
Mother university		-0.261 (1.021)	0.050 (0.771)	0.072 (0.771)	0.025 (0.770)	-0.105 (0.839)
Constant	37.422*** (0.748)	33.764*** (1.725)	30.504*** (1.291)	31.330*** (1.350)	30.340*** (1.289)	10.730*** (1.576)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C4: Determinants of Position Sharing personal beliefs (OLS)

	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se	(6) b/se
Info Social Norm	-0.082 (1.027)	-0.089 (0.979)	0.068 (0.770)	-0.311 (1.115)	0.488 (0.745)	3.347*** (1.154)
Distance Social Norm			0.643*** (0.015)	0.643*** (0.015)	0.695*** (0.022)	
Info Social Norm $\times$ Woman				0.757 (1.543)		
Info Social Norm $\times$ Distance					-0.093*** (0.030)	
Positive distance dummy						30.914*** (1.259)
Info Social Norm $\times$ Positive distance						-5.245*** (1.670)
Absolute distance						0.197*** (0.031)
Woman		-11.965*** (0.980)	-13.263*** (0.767)	-13.685*** (1.170)	-13.214*** (0.766)	-11.878*** (0.838)
Romania		9.257*** (1.307)	12.700*** (1.018)	12.698*** (1.018)	12.745*** (1.014)	10.540*** (1.108)
Sweden		8.584*** (1.218)	11.888*** (0.965)	11.883*** (0.965)	11.859*** (0.964)	10.312*** (1.046)
Religious		8.166*** (1.052)	5.736*** (0.846)	5.730*** (0.847)	5.811*** (0.846)	6.301*** (0.911)
University		-5.193*** (1.070)	-3.062*** (0.859)	-3.067*** (0.860)	-3.088*** (0.857)	-3.872*** (0.923)
Student		-0.815 (1.026)	-1.788** (0.821)	-1.785** (0.821)	-1.800** (0.819)	-1.490* (0.887)
Born abroad		5.922*** (2.053)	3.160* (1.646)	3.174* (1.646)	3.072* (1.634)	3.492** (1.741)
Metropolitan area		-1.022 (1.232)	-0.326 (0.992)	-0.326 (0.992)	-0.396 (0.991)	-0.503 (1.070)
Other/unspecified area		10.334 (9.844)	12.801 (8.046)	12.799 (8.032)	13.246* (7.759)	13.113 (10.860)
Rural area		2.674* (1.673)	2.307* (1.305)	2.302* (1.304)	2.285* (1.304)	1.795 (1.426)
Small city		0.911 (1.293)	0.723 (1.011)	0.707 (1.012)	0.653 (1.010)	0.507 (1.107)
Mother employed		-2.748** (1.149)	-0.794 (0.909)	-0.793 (0.909)	-0.918 (0.908)	-1.526 (0.984)
With partner		1.677* (1.008)	0.518 (0.795)	0.528 (0.796)	0.448 (0.794)	0.877 (0.864)
Mother university		-0.679 (1.057)	-0.310 (0.854)	-0.304 (0.854)	-0.275 (0.850)	0.054 (0.917)
Constant	42.074*** (0.786)	41.052*** (1.815)	36.403*** (1.417)	36.616*** (1.483)	36.280*** (1.413)	18.494*** (1.785)
Observations	3510	3506	3506	3506	3506	3506

Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C5: Determinants of Social Media personal beliefs (OLS)

	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se	(6) b/se	(7) b/se
Info Social Norm (treated)	0.505 (0.728)	0.529 (0.672)	-0.453 (0.903)	-1.409 (0.952)	-0.897 (0.702)	0.783* (0.451)	3.675*** (0.695)
Revenge porn vignette	1.023** (0.483)	1.023** (0.484)	0.243 (0.725)	1.023** (0.484)	1.592*** (0.424)	1.587*** (0.423)	1.438*** (0.446)
Posizione vignette	2.813*** (0.509)	2.815*** (0.510)	1.644** (0.787)	2.815*** (0.510)	2.006*** (0.450)	2.012*** (0.449)	2.231*** (0.475)
Social media vignette	5.112*** (0.512)	5.099*** (0.512)	4.864*** (0.781)	5.099*** (0.512)	6.048*** (0.444)	6.042*** (0.444)	5.280*** (0.475)
Treated × Revenge porn			1.402 (0.973)				
Treated × Posizione			2.107** (1.031)				
Treated × Social media			0.422 (1.035)				
Woman		-12.168*** (0.669)	-12.168*** (0.669)	-14.327*** (1.035)	-16.217*** (0.728)	-14.667*** (0.476)	-12.276*** (0.521)
Treated × Woman				3.875*** (1.345)	2.733*** (0.959)		
Distance to social norm					0.640*** (0.010)	0.670*** (0.015)	
Treated × Distance						-0.055*** (0.020)	
Positive distance dummy							30.453*** (0.748)
Treated × Positive distance							-4.945*** (0.991)
Absolute distance							0.176*** (0.019)
Romania		4.988*** (0.917)	4.988*** (0.917)	4.981*** (0.914)	7.694*** (0.629)	7.708*** (0.630)	6.442*** (0.696)
Sweden		7.193*** (0.842)	7.193*** (0.842)	7.171*** (0.841)	9.700*** (0.617)	9.711*** (0.617)	8.053*** (0.664)
Religious		6.038*** (0.727)	6.038*** (0.727)	6.009*** (0.726)	4.237*** (0.514)	4.281*** (0.514)	4.791*** (0.559)
University		-4.632*** (0.733)	-4.632*** (0.733)	-4.653*** (0.732)	-3.107*** (0.538)	-3.087*** (0.539)	-3.347*** (0.580)
Student		-2.016*** (0.707)	-2.016*** (0.707)	-2.001*** (0.707)	-2.106*** (0.507)	-2.136*** (0.507)	-2.112*** (0.554)
Born abroad		2.782* (1.449)	2.782* (1.450)	2.848** (1.437)	0.735 (1.058)	0.660 (1.060)	1.309 (1.132)
Mother employed		-1.719** (0.794)	-1.719** (0.794)	-1.714** (0.793)	-1.178** (0.566)	-1.243** (0.566)	-1.290** (0.616)
With partner		1.851*** (0.694)	1.851*** (0.694)	1.902*** (0.693)	0.582 (0.498)	0.514 (0.498)	0.859 (0.543)
Mother university		0.775 (0.725)	0.775 (0.725)	0.803 (0.724)	0.703 (0.524)	0.685 (0.523)	0.604 (0.569)
Constant	36.635*** (0.644)	37.603*** (1.296)	38.149*** (1.340)	38.688*** (1.359)	34.944*** (0.981)	34.049*** (0.935)	16.344*** (1.125)
Observations	14040	14024	14024	14024	14024	14024	14024

Robust standard errors clustered at the individual level in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C6: Determinants of Belief Values across Vignettes (OLS, clustered SEs)

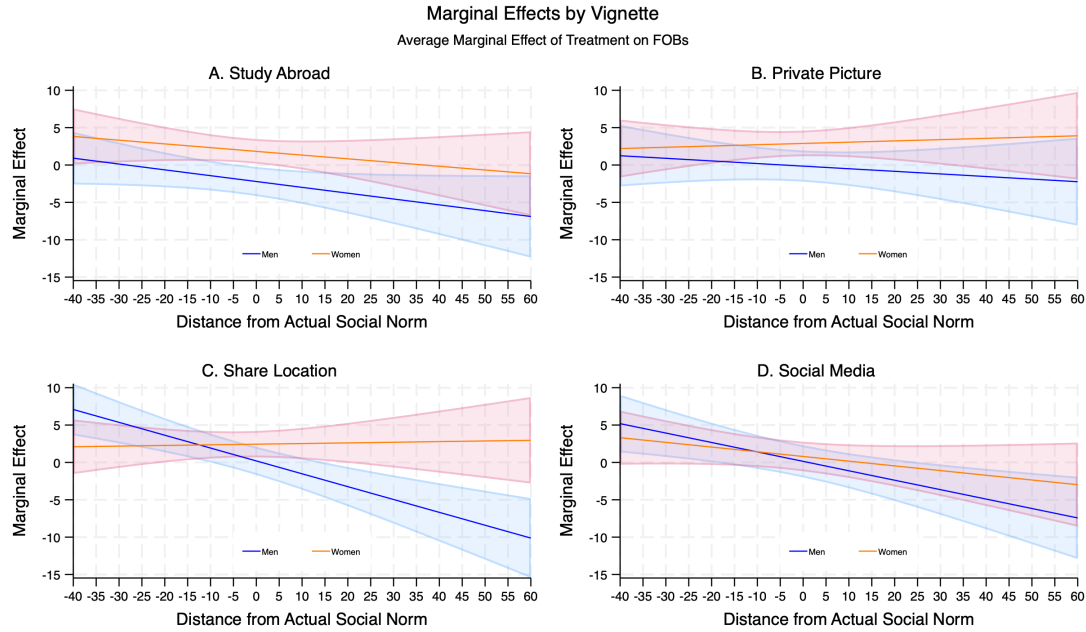


Figure 3: Marginal effects of the treatment on personal beliefs across vignettes by gender (blue line= Man; orange line= Woman). The figure reports average marginal effects of the treatment, with 95% confidence intervals, for each vignette (Study Abroad, Private Picture, Location Sharing, and Social Media).

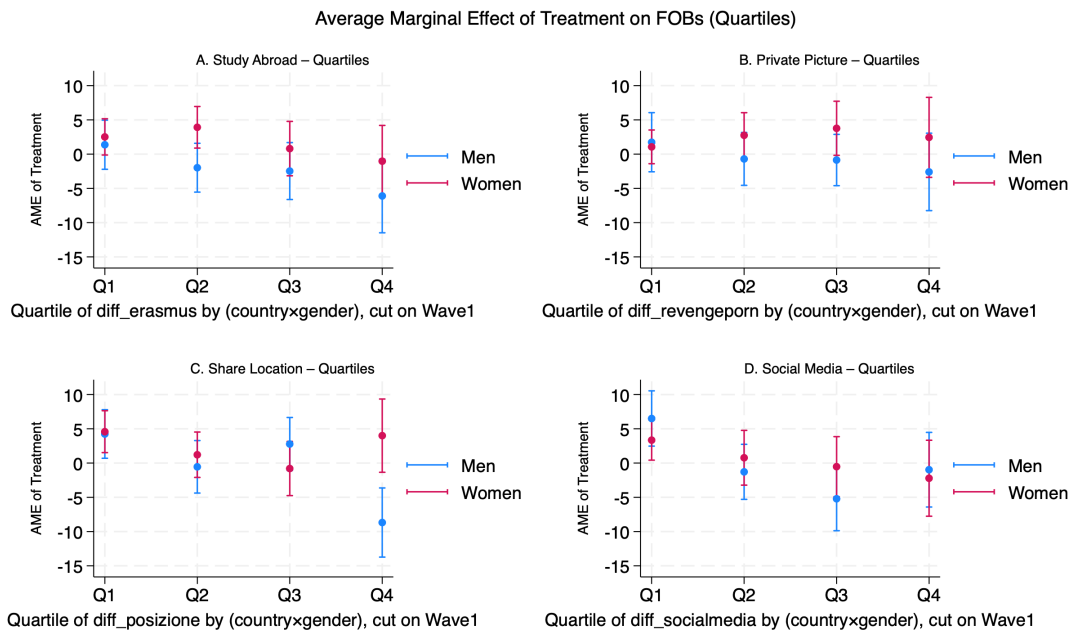


Figure 4: Marginal effects of the treatment on personal beliefs across vignettes by gender, by quartiles of distance to the actual social norm. The figure reports average marginal effects (AMEs), with 95% confidence intervals, by vignette and quartile.

	Dependent variable: Log(Amount Donated)					
	(1)	(2)	(3)	(4)	(5)	(6)
Info Social Norm	0.0329 (0.0441)	0.0350 (0.0443)	0.0355 (0.0445)	0.0505 (0.0642)	0.0186 (0.0469)	0.0027 (0.0845)
Info $\times$ Woman				-0.0321 (0.0879)		
Woman		0.143*** (0.0440)	0.132*** (0.0448)	0.150** (0.0670)	0.127*** (0.0448)	0.136*** (0.0445)
Distance Social Norm_Study Abroad			0.0002 (0.0010)	0.0002 (0.0010)	0.0009 (0.0018)	-0.0014 (0.0867)
Info $\times$ Distance Social Norm_Study Abroad					-0.0014 (0.0022)	-0.0621 (0.1060)
Distance Social Norm_Private Picture			0.0009 (0.0010)	0.0009 (0.0010)	-0.0022 (0.0015)	-0.0781 (0.0751)
Info $\times$ Distance Social Norm_Private Picture					0.0058*** (0.0019)	0.245** (0.0983)
Distance Social Norm_Share Location			0.0008 (0.0010)	0.0008 (0.0010)	0.0012 (0.0017)	0.0555 (0.0800)
Info $\times$ Distance Social Norm_Share Location					-0.0008 (0.0021)	-0.0096 (0.1008)
Distance Social Norm_Social Media			0.0007 (0.0009)	0.0007 (0.0009)	0.0014 (0.0016)	0.1109 (0.0792)
Info $\times$ Distance Social Norm_Social Media					-0.0013 (0.0020)	-0.1007 (0.0997)
Distance_Study Abroad (abs. value)						0.0035** (0.0015)
Distance_Private Picture (abs. value)						-0.0016 (0.0014)
Distance_Share Location (abs. value)						0.0013 (0.0016)
Distance_Social Media (abs. value)						-0.0006 (0.0016)
Observations	1608	1607	1607	1607	1607	1607

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports estimated treatment effects of the *Info Social Norm* intervention on the log of the amount donated to the NGO WAVE. Models progressively include controls and interactions with gender and measures of distance between perceived and actual social norms for each vignette. Columns (1)–(3) use continuous distance measures; Columns (4)–(6) introduce gender interactions and binary distance indicators for positive misperceptions. Controls include country of residence, religiosity, student status, education level, mother’s education and employment, being in a relationship, and whether the respondent was born abroad.

Table C7: Effect of Information Treatment on Log(Amount Donated) on employed respondents

Appendix D

This appendix reports the multiple hypothesis testing (MHT) adjustment applied to the main families of outcomes analysed in the paper. For each family, we compute Romano–Wolf stepdown adjusted p-values using a bootstrap procedure that controls the family wise error rate while accounting for dependence across tests within the same family.

Implementation

We implement the MHT adjustment in Stata using the commands `rwolf` and `rwolf2`. For each family, we estimate the corresponding regression specification and report both the conventional (model) p-value and the Romano–Wolf stepdown adjusted p-value. All adjustments use 5000 bootstrap replications and a fixed seed of 12345. All specifications include the same set of controls as in the main analysis (country fixed effects and baseline socio demographic covariates).

The outcome families are:

- **Family 1:** PV1 (FOB, Option A) treatment main effect across vignettes (4 tests).
- **Family 2:** PV2 treatment main effect across vignettes (4 tests).
- **Family 3:** PV1 gender heterogeneity (treatment $\times$ woman) across vignettes (4 tests).
- **Family 4:** PV2 gender heterogeneity (treatment $\times$ woman) across vignettes (4 tests).
- **Family 5:** PV1 heterogeneity by continuous misperception (treatment $\times$ distance), vignette specific (4 tests).
- **Family 6:** PV2 heterogeneity by continuous misperception (treatment $\times$ distance), vignette specific (4 tests).
- **Family D1:** Donation outcomes, treatment main effect (2 tests: probability of donating and log donation amount).
- **Family D2:** Donation outcomes, gender heterogeneity (2 tests: interaction treatment $\times$ woman).

- **Family P1:** Policy support outcomes, treatment main effect (4 tests).
- **Family P2:** Policy support outcomes, gender heterogeneity (treatment $\times$ woman) (4 tests).

Results and interpretation

Table D1 reports both the conventional p-values from the corresponding regression models and the Romano–Wolf stepdown adjusted p-values computed within each outcome family. The adjustment is performed separately for each family and controls the family wise error rate, while allowing for dependence across tests within the same family. Overall, the Romano–Wolf correction mainly affects results that are only marginally significant under conventional inference. The strongest findings remain statistically significant, while weaker, vignette specific results generally do not survive the stepdown correction.

personal beliefs (PV). For the average treatment effect (Families 1 and 2), none of the PV1 outcomes remains significant after adjustment. In contrast, for PV2 (Family 2) the treatment effect remains significant for *Study Abroad* (adjusted $p = 0.0346$) and *Private Picture* (adjusted $p = 0.0008$), while the remaining two vignettes are not significant (adjusted $p = 0.8772$ in both cases).

For gender heterogeneity (Families 3 and 4), the interaction term *treated* $\times$ *woman* remains significant for PV1 in *Study Abroad* (adjusted $p = 0.0044$) and *Private Picture* (adjusted $p = 0.0376$). The same interaction does not survive the adjustment for *Share Location* (adjusted $p = 0.1636$) and *Social Media* (adjusted $p = 0.3585$). For PV2 gender heterogeneity (Family 4), no interaction effect is statistically significant after adjustment (all adjusted $p \geq 0.3891$).

For heterogeneity by misperception (Families 5 and 6), the continuous interaction *treated* $\times$ *distance* remains significant in PV1 for *Share Location* (adjusted $p = 0.0390$) and *Social Media* (adjusted $p = 0.0094$), while it is not significant for *Study Abroad* (adjusted $p = 0.1254$) and *Private Picture* (adjusted $p = 0.9886$). For PV2, none of the *treated* $\times$ *distance* interactions survives the adjustment (Family 6, all adjusted $p \geq 0.3697$).

Table D1: Multiple testing correction (Romano–Wolf stepdown): model and adjusted p-values

Family	Outcome (tested coefficient)	Model p-value	Romano–Wolf p-value
Family 1: PV1 ATE (treated), 4 tests			
	PV1 Study Abroad (treated)	0.6324	0.8488
	PV1 Private Picture (treated)	0.0568	0.1836
	PV1 Share Location (treated)	0.3336	0.6569
	PV1 Social Media (treated)	0.9278	0.9252
Family 2: PV2 ATE (treated), 4 tests			
	PV2 Study Abroad (treated)	0.0112	0.0346
	PV2 Private Picture (treated)	0.0002	0.0008
	PV2 Share Location (treated)	0.6486	0.8772
	PV2 Social Media (treated)	0.7383	0.8772
Family 3: PV1 gender (treated $\times$ woman), 4 tests			
	PV1 Study Abroad (treated $\times$ woman)	0.0011	0.0044
	PV1 Private Picture (treated $\times$ woman)	0.0125	0.0376
	PV1 Share Location (treated $\times$ woman)	0.0847	0.1636
	PV1 Social Media (treated $\times$ woman)	0.3504	0.3585
Family 4: PV2 gender (treated $\times$ woman), 4 tests			
	PV2 Study Abroad (treated $\times$ woman)	0.1265	0.3891
	PV2 Private Picture (treated $\times$ woman)	0.2727	0.6061
	PV2 Share Location (treated $\times$ woman)	0.8283	0.8324
	PV2 Social Media (treated $\times$ woman)	0.3484	0.6061
Family 5: PV1 heterogeneity by distance (treated $\times$ distance), 4 tests			
	PV1 Study Abroad (treated $\times$ distance)	0.0648	0.1254
	PV1 Private Picture (treated $\times$ distance)	0.9885	0.9886
	PV1 Share Location (treated $\times$ distance)	0.0137	0.0390
	PV1 Social Media (treated $\times$ distance)	0.0020	0.0094
Family 6: PV2 heterogeneity by distance (treated $\times$ distance), 4 tests			
	PV2 Study Abroad (treated $\times$ distance)	0.1110	0.3697
	PV2 Private Picture (treated $\times$ distance)	0.5977	0.5971
	PV2 Share Location (treated $\times$ distance)	0.3434	0.5697
	PV2 Social Media (treated $\times$ distance)	0.1294	0.3697
Family D1: Donation ATE (treated), 2 tests			
	Donate (binary, treated)	0.3746	0.6103
	Log donation (treated)	0.4452	0.6103
Family D2: Donation gender (treated $\times$ woman), 2 tests			
	Donate (binary, treated $\times$ woman)	0.9674	0.9668
	Log donation (treated $\times$ woman)	0.7636	0.9478
Family P1: Policy ATE (treated), 4 tests			
	Sex education (treated)	0.9275	0.9992
	Awareness campaigns (treated)	0.1730	0.4649
	Harsher penalties (treated)	0.8837	0.9992
	Anti violence centres (treated)	0.9184	0.9992
Family P2: Policy gender (treated $\times$ woman), 4 tests			
	Sex education (treated $\times$ woman)	0.9569	0.9564
	Awareness campaigns (treated $\times$ woman)	0.3007	0.4635
	Harsher penalties (treated $\times$ woman)	0.0337	0.1084
	Anti violence centres (treated $\times$ woman)	0.1847	0.3867

Note: The table reports the conventional p-values from the corresponding regression models and the Romano–Wolf stepdown adjusted p-values computed within each family. The adjustment is performed separately for each outcome family.

Donation outcomes. For donation behaviour (Families D1 and D2), neither the treatment main effect nor the gender interaction is statistically significant after adjustment (all adjusted $p \geq 0.6103$).

Policy support outcomes. For policy support (Family P1), there is no evidence of an average treatment effect after adjustment (all adjusted $p \geq 0.4649$). For policy gender heterogeneity (Family P2), the conventional model p-value for *harsher penalties* is below 0.05, but it does not survive the Romano–Wolf correction (adjusted $p = 0.1084$). All other policy interaction tests are also not significant after adjustment.

Moral Panics and the Transformations of Political Cleavages: The STOP-ERA Campaign

Federica Esposito*

Abstract

Can reactionary leaders transform latent social divisions into salient political cleavages by mobilizing moral panics? I study the case of Phyllis Schlafly's STOP-ERA campaign (1972–1982), which played a central role in organizing opposition to the Equal Rights Amendment in the United States. Using a newly constructed dataset on Schlafly's appearances across U.S. counties, I exploit variation in campaign exposure to examine its impact on voting behavior and public opinion. Findings show that local exposure to the campaign increased Republican vote share and shifted beliefs toward more anti-feminist sentiments. Analysis of the mechanism suggests that moral panics can contribute to the transformation of political cleavages by aligning cultural conservatism with partisan identity.

Keywords: moral panics, political cleavages, anti-feminism, historical data, political alignment, cultural conservatism

JEL codes: D72, D83, J16, N32, Z13

*University of Bologna

1 Introduction

A recurrent feature of political conflict in Western democracies is that issues initially characterized by broad consensus or cross-party bargaining can, over time, develop into partisan cleavages. Historical examples include immigration or reproductive rights, each of which gradually evolved into polarizing dimensions of party competition. A growing body of literature focuses on the dynamics of polarization and the transformation of political cleavages. However, the causal impact of political actors in shaping these cleavages remains less well understood.

This paper examines the politicization of women’s rights, a topic that has become one of the most salient and divisive dimensions of contemporary politics. I focus on the United States in the 1970s, a pivotal period for women’s rights. During this decade, the Equal Rights Amendment (ERA)—a constitutional proposal to guarantee equal rights under the law regardless of sex—was introduced with broad bipartisan support and successfully passed Congress. Its ratification process, however, became the subject of intense controversy, leaving a lasting mark on partisan divisions surrounding gender-related policies.

A central figure in this opposition was Phyllis Schlafly, a conservative Republican activist who traveled across the country to mobilize against the ERA. Schlafly framed feminism not merely as a policy dispute but as a broader threat to family stability, motherhood, and Christian values. Her activism can be interpreted as generating a moral panic—a short, intense wave of public alarm about a perceived threat to social values, driven by specific agents (Goode and Ben-Yehuda 1994). In line with moral panic literature, the group portrayed as threatening the social order are *folk demons* (Cohen, 2011). In this perspective, feminists are constructed as the *folk demons* of the STOP-ERA campaign, depicted as threatening family and motherhood values.

Schlafly’s moral panic and associated campaign provide a valuable context to examine the role of reactionary leadership in reshaping political cleavages. Given that party positions on women’s rights and related policies were relatively undefined and of low salience at the beginning of the ERA debate, this setting allows to investigate how such mobilization influences individual beliefs and voting behavior.

I combine county-level election returns with individual-level data from the American National Election Studies (ANES) to investigate both the electoral and attitudinal consequences of Schlafly’s campaign. Using local newspaper archives, I trace Schlafly’s campaign events across the U.S. from 1972 to 1982. I exploit variation in the timing and location of these events to implement an event-study and a difference-in-differences design. Findings show that counties exposed to a Schlafly appearance experienced an increase of roughly 2 percentage points in Republican vote share in House elections. These effects do not persist beyond four months, underscoring the temporally bounded nature of moral

panics.

At the individual level, the analysis of survey data suggests that the voting effects were concentrated among socially conservative voters, who became significantly more likely to support Republican candidates and to view the party as aligned with conservative values. At the same time, exposure to Schlafly’s campaign was associated with a broader increase in hostility toward feminism and a decrease in the support for the ERA across the electorate.

Although the direct electoral effects of Schlafly’s activism were short-lived, the episode reveals how political agents can use moral panics to redefine partisan cleavages. By mobilizing fears about feminism and gender roles, the STOP-ERA campaign helped defining the association between social conservatism and Republican identity—a legacy that continues to shape U.S. politics.

This paper links the study of leadership and mobilization to research on political polarization and cleavage transformation. A growing body of work highlights the role of leaders in shaping political attitudes and electoral outcomes, ranging from the role of leadership in nation-building (Assouad, 2020), to the diffusion of extreme political behaviors (Cagé et al., 2023), the mobilization of social movements (Dippel and Hebllich, 2021), and the effects of leaders’ speeches on voter support (Selb and Munzert, 2018). Similarly, the paper highlights how a leader-driven moral panic can reconfigure partisan conflict around a cultural and social issue—women’s rights—that, prior to the 1970s, was not sharply divided among parties and voters.

Work on polarization and cleavage transformation comprises two complementary strands. First, causal studies focus on possible causes that may have increased polarization (e.g., Autor et al., 2020; Baysan 2022). Second, structural and theoretical analyses clarify the economic and cultural channels and the role of identity and values: Danieli et al. (2022) decompose the rise of the populist radical right into economic and cultural components; Gennaioli and Tabellini 2019, Bonomi et al. 2021 and Enke 2020 emphasize the centrality of identity and moral values; and Gethin et al. 2022 map the long-run reconfiguration of partisan cleavages across Western democracies. In this setting, the causal role of political leaders in driving these transformations remains less clear. I exploit the STOP-ERA campaign as an empirical setting: in the early 1970s in the United States—after the ERA passed Congress—support for women’s rights was broadly bipartisan. Over the subsequent decade, partisan positions sorted, with Republicans increasingly associated with a culturally conservative, pro-life, anti-ERA stance and Democrats remaining pro-ERA. Phyllis Schlafly’s STOP-ERA campaign took place during this transition and generated dated, county-level variation in exposure through local appearances. I leverage this variation to identify causal effects of exposure on perceptions and partisan alignment around women’s rights.

Finally, the paper contributes to the literature on social movements and protest

dynamics. Much of this work has examined progressive mobilizations, showing how collective action can affect political participation, policy outcomes, and institutional change. [Acemoglu et al. 2018] on the Arab Spring or [Melios and Klein Teeselink 2025] on the political effects of Black Lives Matter and [Levy and Mattsson 2023] on the MeToo. [Gethin and Pons 2024] document how mobilization can reshape political preferences. Other work has analyzed conservative movements, such as [Madestam et al. 2013] on the Tea Party, which illustrates how grassroots protest can influence electoral outcomes and policy debates and [Caprettini et al. 2022] focus on the impact of Nazi marches in 1932 Hamburg. My analysis focuses on a reactionary and anti-feminist mobilization—Schlafly’s anti-ERA campaign. By documenting its electoral and attitudinal effects, the paper highlights how movements can also reinforce conservative constituencies and contribute to the politicization of cultural issues.

The remainder of the paper proceeds as follows. Section 2 provides historical background on the ERA and the STOP-ERA campaign. Section 3 describes the data sources on electoral outcomes and public opinion. Section 4 outlines the empirical strategy. Section 5 presents the main results from the county-level event study. Section 6 investigates mechanisms, focusing on partisan perceptions and anti-feminist sentiment. Section 7 concludes.

2 Historical Background

2.1 Gender Politics and Party System before the 1970s

Prior to the 1970s, gender-related issues were not particularly salient or divisive in American partisan politics ([Freeman, 2008]). Both the Democratic and Republican parties included support for equal rights in their platforms. Figure 1 from [Wolbrecht 2010] shows the average pro-women’s rights cosponsorships by each party’s members, represented as the difference between Democratic and Republican cosponsorship. As observed, until the 1970s, there were no significant differences in how the two parties supported women’s rights policies. Additionally, focusing solely on roll call votes, there were only two recorded votes on gender-related issues between 1953 and 1970.

The Equal Rights Amendment (ERA) re-entered the national political debate during this period. Originally proposed in 1923 by Alice Paul and the National Woman’s Party, the ERA aimed to guarantee legal equality between men and women under the Constitution. The Republican Party first endorsed the ERA in 1940, followed by the Democratic Party in 1944. After decades of limited attention, the amendment gained bipartisan support in the early 1970s. In 1972, Congress passed the ERA by overwhelming margins in both chambers and sent it to the states for ratification. Within the first year, 30 of the required 38 states had ratified the amendment.

However, the situation changed in 1980, two years before the ratification deadline. At its 1980 convention, the Republican Party refused to include the ERA in its platform, breaking nearly forty years of official party support. That same year, the Democratic Party not only maintained its pro-ERA stance but also pledged to provide financial support only to candidates who backed the amendment. In 1982, the ERA failed to be ratified. Although it has been reintroduced in Congress multiple times since then, it has never successfully passed both chambers — a situation that remains true to this day.

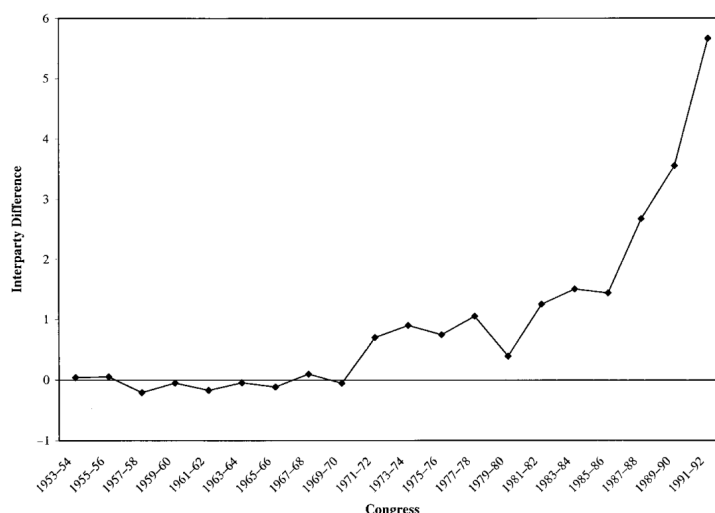


Figure 1: Avg. Pro-Women’s Rights Cosponsorships per each party’s members, House (D - R)

2.2 Phyllis Schlafly and the Moral Panic Against the ERA

During the years of the amendment’s ratification process, a wave of opposition emerged. In particular, the most prominent figure became Phyllis Schlafly. A lawyer, mother of six, and longtime Republican activist, Schlafly launched the STOP-ERA campaign in 1972 to prevent the amendment’s ratification at the state level. Although the ERA initially enjoyed broad bipartisan support and favorable public opinion, Schlafly framed it as a dangerous attack on traditional gender roles. Her rhetoric warned that ratification would lead to women being drafted into combat, the erosion of legal protections like alimony, the abolition of single-sex facilities, and the legalization of abortion and same-sex marriage. In this framing, the name of her campaign was itself emblematic: “STOP” stood for Stop Taking Our Privileges, signaling an effort to mobilize opposition by portraying the ERA not as a vehicle for equality, but as a source of material and moral loss. Through these claims, Schlafly cast feminists as symbolic threats who endangered family and femininity. As evidence of these narratives, there are several issues of her Phyllis Schlafly Report newsletter that present all of her main arguments against the ERA and the feminist movement. The February 1972 issue marks the first time Schlafly addressed the ERA

in her newsletter, as it had not been mentioned in earlier editions. Figure 2 shows two vignettes from that issue: one depicts a sad woman without financial support, and the other shows an injured child returning from war—illustrating Schlafly’s claim that the ERA would have forced women into military service.

All the characteristics of this narrative can be defined as a moral panic. Moral panic is mainly a sociological concept, first defined in the 1970s by Cohen and later by Goode and Ben-Yehuda (1994), who developed a formal model of definition. It can be defined as a situation in which a specific group or social change is perceived as an existential threat to prevailing societal norms. Moral panics are typically characterized by five interrelated elements: widespread concern over a perceived threat, increased hostility toward the group associated with it, a consensus (even if not universal) that the threat is real and urgent, a disproportionate reaction relative to the actual risk, and a volatile rise and fall in public attention and emotional intensity (Goode and Ben-Yehuda, 1994).

Within this framework, the STOP ERA campaign led by Phyllis Schlafly in the 1970s satisfies the criteria of a moral panic. Schlafly’s narrative generated concern and mobilization through speeches, media appearances, and grassroots activism. According to some scholars, Schlafly’s campaign was fundamental to the defeat of the ERA (Critchlow, 2005; Spruill, 2017), with some arguing that without her, the amendment would have been ratified (Mansbridge, 2015).

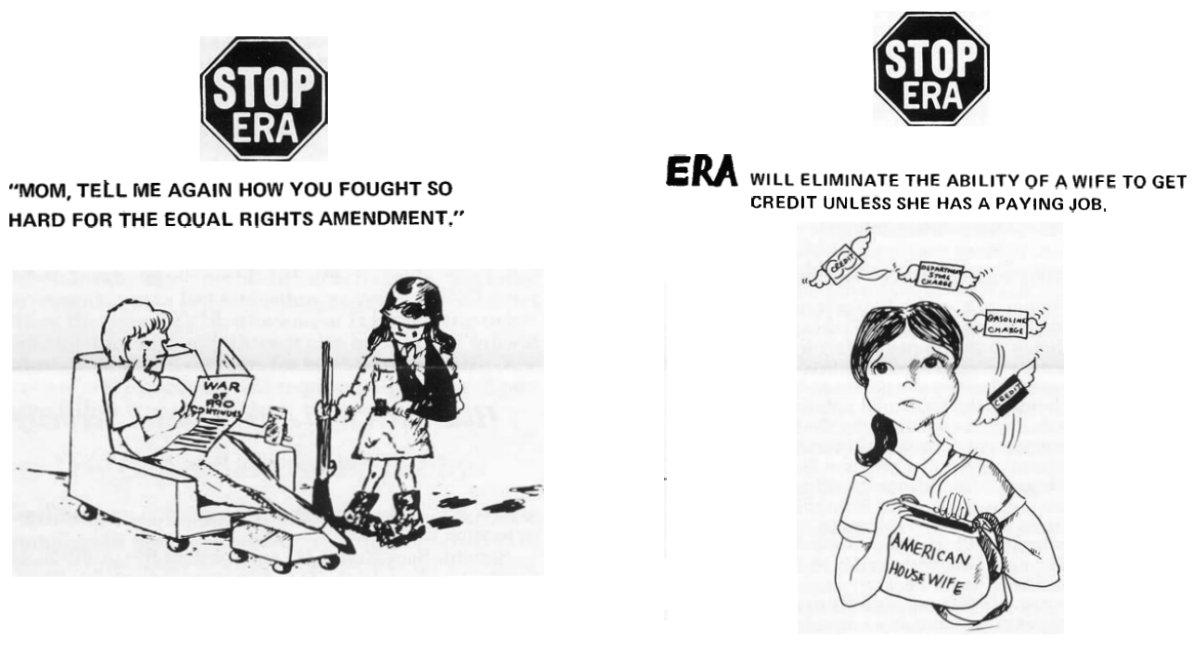


Figure 2: Anti-ERA illustrations from the *Phyllis Schlafly Report*.

3 Data

3.1 Schlafly’s Campaign Appearances

To track the campaign appearances of Phyllis Schlafly during the STOP-ERA campaign (1972-1982), a novel dataset of her public events was constructed using 42617 digitized newspaper pages. The source of the data is Newspapers.com, one of the largest historical newspaper archives, from which all pages mentioning “Phyllis Schlafly” were retrieved. Figure 3 shows the mean daily newspaper mention share, computed as the share of newspapers in a county mentioning Schlafly on a given day (and set to 0 on days with no mentions). The figure reports the average of this daily measure over the 1972–1982 period. The share is spread across the entire United States, with non-zero values appearing in many counties in every major region. While the intensity varies—some counties show higher average shares than others—the overall pattern indicates nationwide visibility rather than attention confined to a single area.

To identify the campaign events, duplicate articles—often reprints or syndicated versions appearing across different outlets or editions—were then removed by measuring textual similarity and excluding near-identical entries.¹ Following de-duplication, the dataset was filtered to identify only those articles referring to *actual in-person events*, rather than general mentions of Schlafly. This was accomplished through keyword detection, retaining only articles that included terms typically associated with physical appearances, such as “speak,” “attend,” “meeting,” or “rally.”² The resulting subset comprised articles explicitly describing Schlafly’s participation in specific campaign events. These articles were then manually reviewed to verify their content and extract relevant details. The final dataset is structured so that each entry corresponds to a verified campaign appearance by Schlafly, including information on the location, date of the event, newspaper name and publication date, and type of appearance—such as a rally, university speech or conference, charity event, or debate. The final dataset comprises 339 campaign appearances across 200 U.S. counties. Table 1 shows the distribution of events by type. It reveals that the majority are local events—such as charity, association, or fundraising events while the debate represents the 20% and the rallies the 19.5 %.

Figure 4 shows the counties in which events from Schlafly’s STOP-ERA campaign took place. As is evident, the campaign was geographically widespread and reached across much of the country.

Some notable events included the protests at the public hearing at the state capitol

¹Duplicate removal was based on cosine similarity of TF-IDF vectors. Articles with a similarity score above 0.9 were flagged as redundant, and only one version was retained.

²Filtering was performed using regular expressions to detect both verb structures and event-related nouns (e.g., “conference,” “visit,” “speaker”). A secondary filter, based on natural language processing (SpaCy), was applied to confirm that Schlafly was syntactically the subject or object of action verbs indicating physical presence.

building, as well as the more iconic demonstrations in which Schlafly organized supporters to bring freshly baked bread to state legislators—intended as a symbolic gesture of women’s willingness to embrace traditional domestic roles (see Figures A1 in Appendix A).

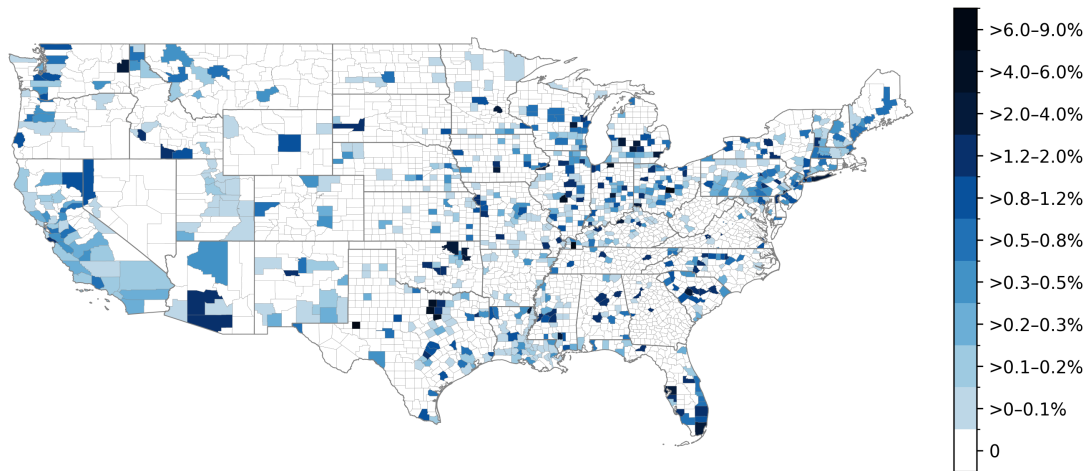


Figure 3: Mean daily newspaper mention share of Phyllis Schlafly, 1972–1982 (by county).

Event type	Share
Charity / associations / fundraisers	41.29%
Debate	20.98%
Campaign / rally / protest	19.20%
State building hearing	9.60%
Republican conventions / party events	8.93%

Table 1: Composition of Phyllis Schlafly events by type

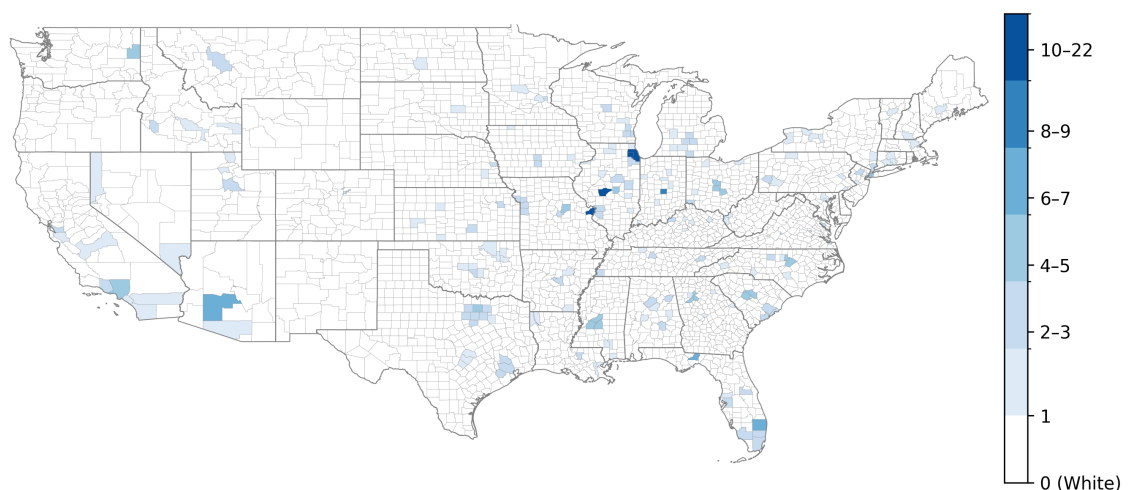


Figure 4: Distribution of Phyllis Schlafly campaign events by county

3.2 Public Opinion

To identify the effect of the STOP-ERA campaign and the associated moral panic on public opinion, I focus primarily on voting behavior. Given data availability, my analysis

is conducted at the county level. I use the ICPSR dataset on congressional election returns (1950–1990), focusing on U.S. House of Representatives elections, which occur biennially. For each election, I observe county-level vote counts by party, as well as total votes cast. Concentrating on the Republican vote share, I examine whether exposure to Phyllis Schlafly’s anti-feminist rhetoric led to a shift in support for the Republican Party.

To gain insight into who is most persuaded by Schlafly’s narrative and how the campaign may have shaped individual beliefs, I complement this with data from the American National Election Studies (ANES), specifically the Cumulative Time Series (1948–2020). This dataset allows me to track public attitudes on women’s issues—such as opposition to abortion, and support for feminist movements—alongside politically salient variables like self-reported voting and perceptions of the parties positions (e.g., whether they are viewed as more or less conservative).

4 Empirical Strategy

To identify the effects of exposure to the STOP-ERA campaign on political behavior, I exploit the variation in Schlafly’s public appearances across counties and over time. Specifically, I use an event-study framework to estimate the dynamic effect of local campaign activity on Republican vote share in US House elections, which occur biennially. The main specification is given by:

$$y_{ct} = \gamma \cdot D_{ct}(-\infty, -601) + \sum_{\ell=-600}^{-121} \beta_{\ell} D_{ct}(\ell, \ell+120) + \sum_{\ell=1}^{600} \beta_{\ell} D_{ct}(\ell, \ell+120) + \delta \cdot D_{ct}(601, \infty) + \alpha_c + \theta_t + \varepsilon_{ct} \quad (1)$$

where y_{ct} is the Republican vote share in county c and election year t . The sample includes U.S. House elections from 1970 to 1984, which occur every two years.

$D_{ct}(\ell, \ell + 120)$ is an indicator equal to one if a Schlafly event occurred in county c between ℓ and $\ell + 120$ days relative to the election date. When $\ell \in [-600, -121]$, the dummy equals one if the election occurred after the speech (post-treatment windows). When $\ell \in [1, 600]$, the dummy equals one if the election occurred before the speech (pre-treatment windows), which allows testing for pre-trends.

The coefficient γ captures the effect of events that occurred more than 600 days before the election, while δ captures the effect of events that occurred more than 600 days after the election. These coefficients serve as *long-run* leads and lags. County (α_c) and year (θ_t) fixed effects are included to account for time-invariant local characteristics and common time shocks. Standard errors are clustered at the county level. This setup allows me to test for both pre-treatment trends and dynamic post-treatment effects.

To explore the mechanisms underlying these electoral shifts, I complement the county-

level analysis with individual-level data from the American National Election Studies (ANES) Cumulative Time Series (1948–2020). I adopt a difference-in-differences strategy, where treatment is defined at the respondent level based on whether the individual was interviewed shortly after a Schlafly campaign event in their county of residence. The treatment window is chosen to align with the timing of effects identified in the event-study analysis and will be discussed in detail below. I estimate the following model:

$$Y_{ict} = \alpha + \beta D_{ict} + X'_{ict}\gamma + \lambda_c + \delta_t + \theta_{t \times c} + \varepsilon_{ict} \quad (2)$$

where Y_{ict} denotes the political beliefs or attitudes of individual i in county c and year t , and D_{ict} is a treatment indicator. The vector X_{ict} includes demographic controls. This framework allows me to assess whether the STOP-ERA campaign influenced beliefs, party identification, and perceptions of the Republican party.

5 Main Results

5.1 Voting Behavior

Figure 5 presents coefficient estimates from the event-study specification (Equation 1). The regression is weighted by county population to account for differences in voter size and captures the dynamic effect of local exposure to Schlafly’s STOP-ERA campaign on Republican vote share in congressional elections³. Time is divided into 120-day intervals, covering roughly three years before and after the campaign event, which allows for a flexible assessment of the treatment timing.

The estimates reveal a substantial and statistically significant increase in Republican vote share in the months preceding an election. In particular, counties exposed to campaign activity within 1–120 days of an election experienced an increase of 2 percentage points compared to counties not affected by such activity. This effect persists, though with slightly lower significance, in the 121–240 day window, where the point estimate remains at 2 percentage points and is statistically significant at the 90% confidence level. These results indicate that Schlafly’s campaign was successful in mobilizing Republican support when a campaigning event took place within four months before the elections.

Estimates for longer lags—starting from 241 days and beyond—decline sharply and lose statistical significance. This pattern suggests that, while the campaign was effective in generating short- to medium-term partisan shifts, it did not produce enduring realignment across election cycles, which, in a way, confirms the temporary nature of the moral panic’s effects. Importantly, estimates for all pre-treatment periods are close to zero and not

³I exclude Alaska and Hawaii because they are not directly comparable to the contiguous U.S. states: they are geographically isolated, exhibit distinct political and demographic dynamics, and Schlafly did not hold any events there.

statistically significant, reinforcing the credibility of the research design and supporting the parallel trends assumption. As a robustness, in Appendix B, I present the results from the estimator proposed by [De Chaisemartin and d’Haultfoeuille \(2024\)](#) in Figure [B1](#). I also re-estimate the main specification excluding the states that Schlafly never visited (Delaware, Oregon, New Mexico, Rhode Island, and Wyoming), and the results remain similar (Figure [B2](#)). Finally, Table [B1](#) presents the results of two additional specifications of the DiD model: one where treatment is defined by a dummy variable equal to one if at least one campaign event occurred within 120 days prior to the election, and another where treatment is captured by a continuous variable measuring the number of events held in that window. The findings are again consistent. Additionally, I conduct a preliminary heterogeneity analysis by event type. I classify events according to whether they were debates (i.e., featured opposing arguments) or non-debate appearances, and whether Schlafly was the organizer/host or an invited guest. The results are reported in Appendix Table [B2](#). They suggest that events without an opposing argument were more effective, while there is no marked difference between events she organized and those where she was invited. The findings portray the STOP-ERA campaign as a politically consequential, yet temporally bounded, intervention. Schlafly’s events operated as localized political shocks that amplified Republican support in the months preceding an election, but their impact faded relatively quickly, consistent with the episodic and emotional nature of moral panics.

Although the effects are confined to the months surrounding the election, they highlight how campaign-driven salience can temporarily align cultural issues—such as gender roles—with partisan choice. In this sense, the STOP-ERA campaign helped reinforce the emerging association between social conservatism and Republican voting, a key step in the broader transformation of political cleavages.

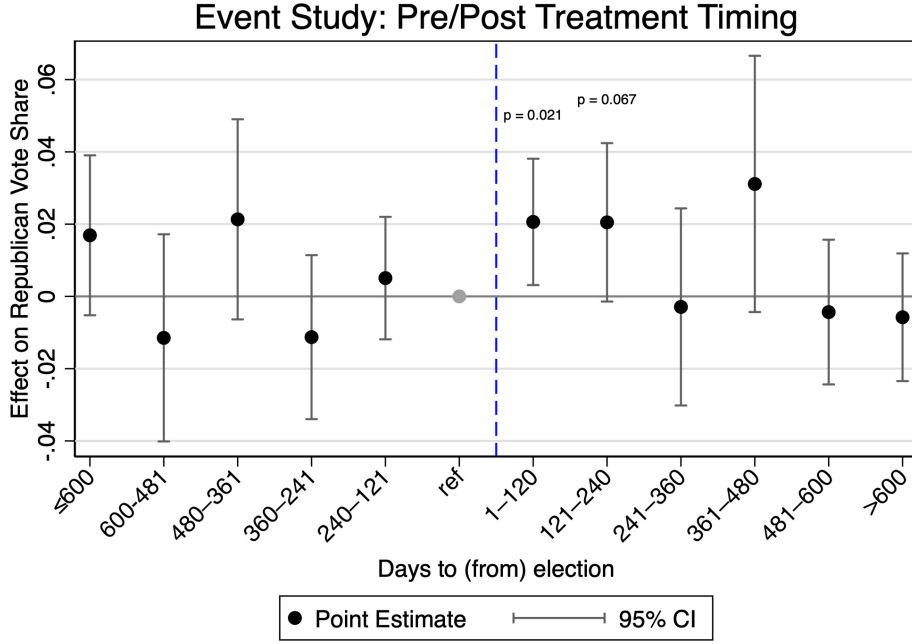


Figure 5: Event-study estimates of Schlafly’s campaign effect on Republican vote share. *Notes:* Number of Observation: 22,511, Weighted by county population in 1980

5.2 Individual Beliefs

In the previous section, I showed that Schlafly’s campaign events increased the Republican Party’s vote share in congressional elections, but only when the events occurred sufficiently close to the election date (approximately within four months). In this section, I turn to the effect of the campaign on individual beliefs concerning the issues emphasized during Schlafly’s public appearances—particularly those related to the ERA, gender equality, and conservative gender norms.

To investigate this, I use data from the American National Election Studies (ANES) and apply the identification strategy outlined in Section 4. I focus on four key outcomes drawn from survey questions: two relate to the political dimension of gender equality—support for the ERA and attitudes toward the feminist movement (Table 2)—while the other two reflect broader underlying beliefs, namely conservative gender norms (as measured by the Women’s Role Scale) and views on abortion rights (Table 3).

The results suggest that Schlafly’s campaign significantly influenced attitudes toward the political dimension of gender equality but had no significant effect on deeper ideological beliefs. In Table 2, I find that exposure to the campaign reduced the probability that a respondent supports the ERA by about 26 percent. At the same time, it increased the probability of expressing cold feelings toward feminists by approximately 10 percent. Support for the ERA is measured using a binary indicator equal to 1 if the respondent agrees or strongly agrees with the amendment. Cold feelings toward feminists are captured

using a thermometer scale rescaled into a binary variable, equal to 1 if the respondent gives a score below 50.

By contrast, Table 3 shows no significant effect of the campaign on either conservative gender beliefs or abortion attitudes. Conservative gender norms are measured using the Women’s Role Scale, which ranges from 1 (support for full gender equality) to 7 (agreement that a woman’s place is in the home). Opposition to abortion is coded as 1 if the respondent believes abortion should be prohibited or permitted only when the mother’s life is at risk.

These results suggest that Schlafly’s campaign did not shift respondents’ underlying ideological views, but rather their political attitudes and behavior. In other words, the campaign appears to have activated pre-existing conservative beliefs, making individuals more likely to align with the Republican Party, rather than changing their core values. In the next section, I explore this mechanism in greater detail.

Table 2: Effect of Schlafly’s Campaign on Support for ERA and Attitudes Toward Feminists

	Support for ERA		Cold Feeling toward Feminists	
	(1)	(2)	(3)	(4)
Schlafly’s Campaign	-0.259*** (0.044)	-0.260*** (0.045)	0.087** (0.042)	0.098** (0.041)
Controls	No	Yes	No	Yes
Clusters	163	163	175	175
Observations	5,550	5,471	11,849	11,665

Notes: All models include county, year, and county $\times$ year fixed effects. Robust standard errors clustered at the county level are shown in parentheses. **Cold Feeling toward Feminists** is equal to 1 if in the thermometer series question over Feminists the respondents indicate a values below 50 on a 1-100 scale. Controls include marital status, religion, working status, parental status, church attendance, gender, education, race, income, and age group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Effect of Schlafly’s Campaign on Women’s Roles and Abortion Attitudes

	Women’s Role Scale		Opposition to Abortion	
	(1)	(2)	(3)	(4)
Schlafly’s Campaign	0.066 (0.200)	-0.073 (0.198)	0.077 (0.080)	0.060 (0.071)
Controls	No	Yes	No	Yes
Clusters	175	175	163	163
Observations	10,474	10,333	6,655	6,562

Notes: All models include county, year, and county $\times$ year fixed effects. Robust standard errors clustered at the county level are shown in parentheses. The **Women’s Role Scale** ranges from 1 (men and women should have equal roles in society) to 7 (a woman’s place is in the home). **Opposition to Abortion** equals 1 if the respondent is completely opposed to abortion or allows it only when the mother’s life is in danger. Controls include marital status, religion, working status, parental status, church attendance, gender, education, race, income, and age group.

6 Mechanism

The results presented so far suggest that Schlafly’s campaign influenced voting behavior without altering individuals’ underlying values. Instead, the campaign appears to have affected attitudes related to the political expression of gender equality. To better understand these findings, it is important to identify which voters were most responsive to Schlafly’s message.

To explore this, I examine heterogeneity in the treatment effect along three dimensions drawn from the ANES survey. First, I consider whether the respondent identified as a housewife, a group that was both the target audience of Schlafly’s STOP ERA campaign and actively involved in its mobilization. Second, I use respondents’ positions on abortion rights, based on a question asking under which circumstances abortion should be allowed by law. Third, I employ the Women’s Role Scale, a 1–7 scale measuring support for traditional gender roles, with higher values indicating more conservative views.

These measures help to distinguish respondents more likely to be aligned with Schlafly’s narrative and thus more susceptible to persuasion by her campaign. To implement this heterogeneity analysis, I restrict the sample to individuals who were interviewed both before and after a Schlafly speech in their area, leveraging the pre/post design of the ANES. As in the main analysis, respondents are classified as treated if they were interviewed within 120 days of a campaign event in their county.

Importantly, the variables used to construct the measures of conservative gender norms and opposition to abortion were only asked in the pre-election wave of the 1972 ANES survey within the 1972–1982 period. As a result, the heterogeneity analysis that uses these variables must be restricted to cross-sectional data from 1972. Table 4

explores heterogeneity in the effect of Schlafly’s campaign on Republican vote choice in congressional elections, focusing on individual beliefs and social characteristics linked to gender conservatism. The baseline specification in column (1) indicates that respondents interviewed within four months of a Schlafly event were 25 percentage points more likely to vote Republican, relative to those not exposed. This effect remains strong and statistically significant after the inclusion of individual-level controls (column 2).

Column (3) investigates gender differences and finds no significant difference in the campaign effect between men and women. However, column (4) suggests that housewives—a group directly targeted by Schlafly’s STOP ERA campaign—were more responsive: the interaction term is positive and marginally significant. This pattern is even more pronounced in the women-only sample (column 7), where treated housewives are 8.6 percentage points more likely to vote Republican than other women.

Stronger evidence of heterogeneity emerges in columns (5) and (6), which examine ideological alignment with Schlafly’s message. Column (5) shows that the treatment effect is significantly larger among respondents who oppose abortion rights. For these individuals, exposure to Schlafly’s campaign increased Republican vote choice by an additional 14.6 percentage points, even though abortion opposition on its own is not significantly associated with voting Republican. Column (6) examines the Women’s Role Scale, which ranges from 1 (“Men and women should have equal roles in society”) to 7 (“A woman’s place is in the home”). Respondents scoring 4 or higher on this scale are classified as having conservative gender norms. Among them, the interaction with treatment is strongly positive and statistically significant. This suggests that Schlafly’s campaign specifically mobilized individuals who already held socially conservative views, reinforcing their political alignment with the Republican Party.

Further evidence in support of this mobilization mechanism comes from Table 5, which shows how exposure to Schlafly’s campaign shaped individuals’ perceptions of the Republican Party. Specifically, the dependent variable is an indicator equal to one if the respondent identified the Republican Party as the most conservative among the major parties.

The results indicate that Schlafly’s campaign significantly increased the probability of perceiving the Republican Party as most conservative—by approximately 27 to 31 percentage points across specifications (columns 1–4 and 7). This effect is not driven by gender alone, as the interaction with gender is small and not significant, but becomes particularly pronounced among subgroups ideologically aligned with Schlafly’s message. In particular, the interaction terms in columns (5) and (6) show that the effect is significantly stronger among individuals who oppose abortion or hold conservative gender beliefs.

Schlafly’s campaign mobilized voters who already held socially conservative views. Individuals who endorsed traditional gender roles or opposed abortion were significantly more responsive to her message—both in their voting behavior and in how they perceived

the Republican Party’s ideological position. In this sense, the campaign reinforced and politicized existing social identities, contributing to a stronger alignment between cultural conservatism and partisan affiliation.

These patterns illustrate how targeted campaigns can convert social attitudes into electoral outcomes, particularly when they engage issues that already hold moral or cultural meaning to parts of the electorate.

Table 4: Heterogeneity of Campaign effect with respect to people’s beliefs on gender equality

	Republican Vote in Congressional Elections						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Schlafly’s Campaign	0.250*** (0.037)	0.194*** (0.071)	0.200* (0.108)	0.187** (0.078)	0.107** (0.045)	0.017 (0.072)	0.166*** (0.063)
Woman			0.003 (0.017)				0.000 (.)
Schlafly’s Campaign $\times$ Woman			−0.071 (0.054)				
Housewife				0.043 (0.027)			0.053 (0.036)
Schlafly’s Campaign $\times$ Housewife				0.050* (0.029)			0.086** (0.041)
Conservative Gender Norms						0.005 (0.017)	
Schlafly’s Campaign $\times$ Conservative Gender Norms						0.416*** (0.036)	
Oppose Abortion					−0.021 (0.018)		
Schlafly’s Campaign $\times$ Oppose Abortion					0.146** (0.057)		
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	All	All	Women only
Observations	2,657	2,625	2,620	2,625	2,543	2,625	1,485

Notes: Dependent variable equals 1 if respondent voted Republican in congressional elections. Standard errors clustered at the county level. Controls include: marital status, religion, working status, parental status, church attendance, gender, education, race, income, age group. Conservative Gender Norms = 1 if respondent’s Women’s Role Scale (1–7) $\geq$ 4. “Oppose Abortion” = 1 if the respondent is completely opposed or only supports abortion when the mother’s life is in danger. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: Heterogeneity of Campaign effect with respect to people's beliefs on gender equality

	Perception: Republican Party is Most Conservative						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Schlaflly's Campaign	0.311*** (0.035)	0.275*** (0.044)	0.238*** (0.027)	0.272*** (0.046)	-0.031* (0.017)	0.005 (0.017)	0.304*** (0.064)
Woman			-0.008 (0.020)				
Schlaflly's Campaign $\times$ Woman			0.005 (0.056)				
Housewife				0.052** (0.024)			0.054 (0.036)
Schlaflly's Campaign $\times$ Housewife				0.053* (0.029)			0.071* (0.040)
Conservative Gender Norms						0.005 (0.017)	
Schlaflly's Campaign $\times$ Conservative Gender Norms						0.261*** (0.071)	
Oppose Abortion					-0.031* (0.017)		
Schlaflly's Campaign $\times$ Oppose Abortion					0.257** (0.103)		
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	All	All	Women only
Obs.	2,657	2,625	2,620	2,625	2,543	2,625	1,485

Notes: Dependent variable equals 1 if the respondent reports that the Republican Party is the most conservative. "Conservative Gender Norms" equals 1 if the respondent scores ≥ 4 on the Women's Role Scale (1–7). "Oppose Abortion" equals 1 if the respondent is completely opposed to abortion or allows it only when the mother's life is at risk. All models include county fixed effects. Standard errors clustered at the county level. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

7 Conclusion

This study provides new evidence on how reactionary political leaders can contribute to the formation and consolidation of partisan cleavages by mobilizing moral panics around cultural issues. Using newly collected data on the timing and location of Phyllis Schlafly’s STOP-ERA campaign events between 1972 and 1982, I show that local exposure to the campaign significantly increased Republican vote share in congressional elections. This effect is concentrated in the four months following an event, consistent with the mobilizing impact of a short-term, issue-specific campaign.

To understand the mechanism behind this electoral shift, I complement the event-study analysis with individual-level survey data from the American National Election Studies (ANES). The evidence suggests that Schlafly’s campaign worked not by persuading undecided voters or shifting general ideological orientations, but by politically mobilizing constituencies who already held socially conservative beliefs—particularly those opposed to abortion or supportive of traditional gender roles. These voters were more likely to increase support for the Republican Party after campaign exposure, as shown both in their reported voting behavior and in their perceptions of the Republican Party’s ideological positioning.

Crucially, the campaign also shifted attitudes on specific gender-related political issues. Schlafly’s speeches led to a decrease in support for the Equal Rights Amendment and an increase in cold feelings toward feminists, indicating that the campaign influenced how respondents evaluated the feminist movement and gender equality in political terms. However, the campaign did not change underlying beliefs about women’s roles in society or abortion rights, suggesting that the campaign operated by reshaping the political salience of cultural issues—turning them into partisan signals—rather than altering individual moral beliefs.

Schlafly’s campaign played a key role in reinforcing and politicizing existing social divisions. By linking moral and cultural concerns to partisan identity, the campaign helped consolidate a Republican alignment among socially conservative voters. The STOP-ERA mobilization contributed to the reframing of gender politics as a partisan cleavage.

More broadly, this study contributes to the literature on the transformation of political cleavages and polarization. It suggests that durable cleavages may emerge not only from large-scale socioeconomic transformations but also from the ability of political entrepreneurs to generate moral narratives that are institutionalized through parties and activist organizations. Understanding the conditions under which such narratives become embedded in partisan competition remains crucial for explaining both historical realignments and contemporary identity-based polarization.

References

- D. Acemoglu, T. A. Hassan, and A. Tahoun. The power of the street: Evidence from egypt’s arab spring. *The Review of Financial Studies*, 31(1):1–42, 2018.
- L. Assouad. Charismatic leaders and nation building. 2020.
- D. Autor, D. Dorn, G. Hanson, and K. Majlesi. Importing political polarization? the electoral consequences of rising trade exposure. *American Economic Review*, 110(10): 3139–3183, 2020.
- C. Baysan. Persistent polarizing effects of persuasion: Experimental evidence from turkey. *American Economic Review*, 112(11):3528–3546, 2022.
- G. Bonomi, N. Gennaioli, and G. Tabellini. Identity, beliefs, and political conflict. *The Quarterly Journal of Economics*, 136(4):2371–2411, 2021.
- J. Cagé, A. Dagorret, P. Grosjean, and S. Jha. Heroes and villains: The effects of heroism on autocratic values and nazi collaboration in france. *American Economic Review*, 113(7):1888–1932, 2023.
- B. Caprettini, M. Caesmann, H.-J. Voth, and D. Yanagizawa-Drott. Going viral: Protests and polarization in 1932 hamburg, 2022.
- S. Cohen. *Folk devils and moral panics*. Routledge, 2011.
- D. T. Critchlow. *Phyllis Schlafly and grassroots conservatism: A woman’s crusade*. Princeton University Press, 2005.
- O. Danieli, N. Gidron, S. Kikuchi, and R. Levy. Decomposing the rise of the populist radical right. *Available at SSRN 4255937*, 2022.
- C. De Chaisemartin and X. d’Haultfoeuille. Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, pages 1–45, 2024.
- C. Dippel and S. Heblich. Leadership in social movements: Evidence from the “forty-eighters” in the civil war. *American Economic Review*, 111(2):472–505, 2021.
- B. Enke. Moral values and voting. *Journal of Political Economy*, 128(10):3679–3729, 2020.
- J. Freeman. *We will be heard: Women’s struggles for political power in the United States*. Bloomsbury Publishing PLC, 2008.
- N. Gennaioli and G. Tabellini. Identity, beliefs, and political conflict. Technical report, CESifo Working Paper, 2019.

- A. Gethin and V. Pons. Social movements and public opinion in the united states. Technical report, National Bureau of Economic Research, 2024.
- A. Gethin, C. Martínez-Toledano, and T. Piketty. Brahmin left versus merchant right: Changing political cleavages in 21 western democracies, 1948–2020. *The Quarterly Journal of Economics*, 137(1):1–48, 2022.
- E. Goode and N. Ben-Yehuda. Moral panics: Culture, politics, and social construction. *Annual review of sociology*, 20(1):149–171, 1994.
- R. Levy and M. Mattsson. The effects of social movements: Evidence from# metoo. Available at SSRN 3496903, 2023.
- A. Madestam, D. Shoag, S. Veuger, and D. Yanagizawa-Drott. Do political protests matter? evidence from the tea party movement. *The Quarterly Journal of Economics*, 128(4):1633–1685, 2013.
- J. J. Mansbridge. *Why we lost the ERA*. University of Chicago Press, 2015.
- G. Melios and B. Klein Teeselink. Weather to protest: the effect of black lives matter protests on the 2020 presidential election. *Political Behavior*, 2025.
- P. Selb and S. Munzert. Examining a most likely case for strong campaign effects: Hitler’s speeches and the rise of the nazi party, 1927–1933. *American Political Science Review*, 112(4):1050–1066, 2018.
- M. J. Spruill. *Divided we stand: The battle over women’s rights and family values that polarized American politics*. Bloomsbury Publishing USA, 2017.
- C. Wolbrecht. The politics of women’s rights: Parties, positions, and change. In *The Politics of Women’s Rights*. Princeton University Press, 2010.

Appendix A: Additional Figures



Figure A1: Phyllis Schlafly and STOP-ERA supporters bringing homemade bread to legislators. The Argus (Rock Island Ill.) 19-06-1978

Appendix B

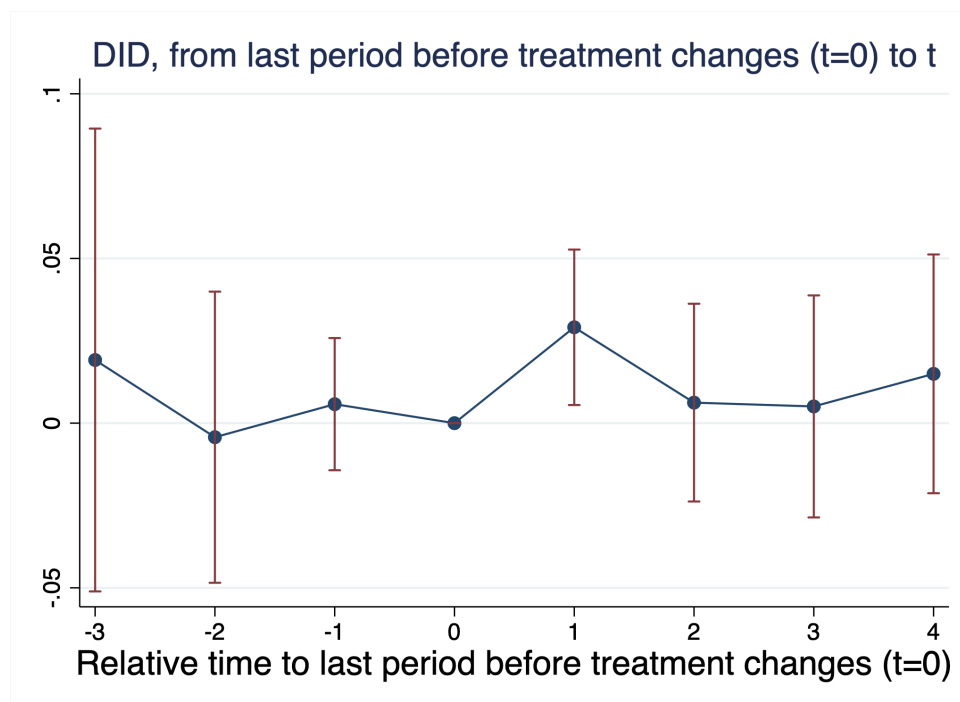


Figure B1: De Chaisemartin and d'Haultfoeuille 2024 Event-Study Estimates.

Table B1: Effect of Speeches on Republican Vote Share (County FE; Year FE; SE clustered by county)

Last 120 Days → Election Day	(5)	(6)	(7)	(8)
Schlafl'y's Campaign	0.0197** (0.0097)		0.0402** (0.0177)	
# Schlafl'y's Campaign Events		0.0207** (0.0096)		0.0402*** (0.0155)
Obs.	22,511	22,511	22,511	22,511
Clusters	3,008	3,008	3,008	3,008
County FE / Year FE	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes
Weights (County Population)	Yes	Yes	No	No

Notes: Outcome is *sharecongreprevotes*. TWFE regressions with county and year fixed effects; standard errors clustered at county level (in parentheses). Sample: 1970–1986, vote shares in (0,1). “Schlafl'y's Campaign” is a dummy equal to 1 if at least one event occurred within 120 days before the election. “# Schlafl'y's Campaign Events” is the count of events in that window. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

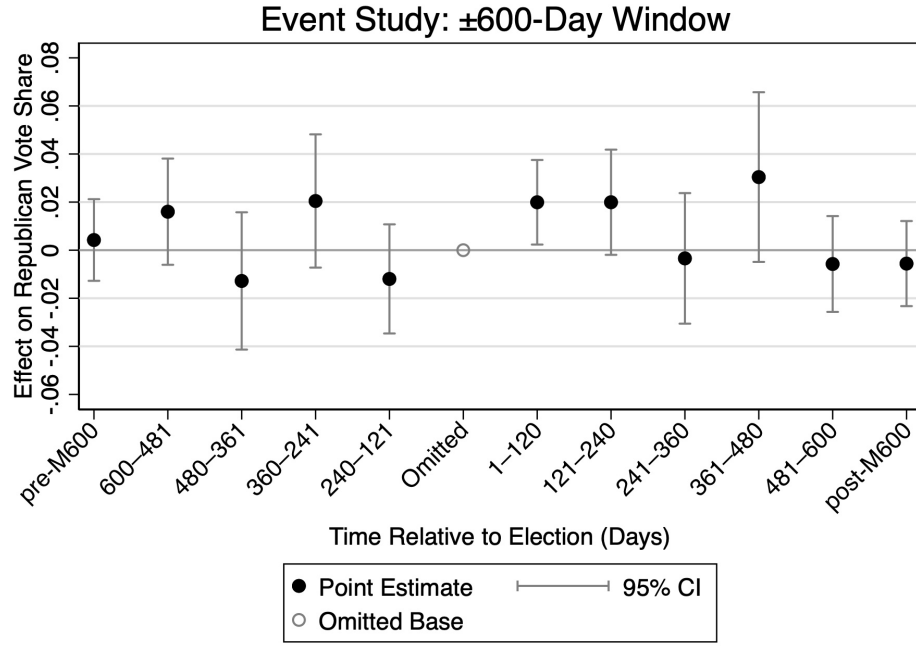


Figure B2: Event-Study Estimates excluding *never treated states*.
Notes: 21,676 Observations, Weighted by county population in 1980

	Panel A: Event format		Panel B: Role in event	
	(1)	(2)	(3)	(4)
No opposing side	0.021** (0.010)	0.042** (0.019)		
Debate	-0.059 (0.045)	-0.011 (0.045)		
Organizer			0.053*** (0.014)	0.045** (0.020)
Guest			0.017* (0.009)	0.033** (0.020)
County population weights	Yes	No	Yes	No
Observations	22,511	22,511	22,511	22,511

Notes: Two-way fixed effects (county and year). Standard errors clustered by county. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: Heterogeneity by Event Type: Republican Vote Share