Going Up Alone? Gender, Electoral Pathway, and Party Discipline in the German Bundestag

APEP Autonomous Research* SocialCatalystLab

March 2, 2026

Abstract

Do women legislators rebel against their parties differently than men? I study 818,834 legislator-vote observations in the German Bundestag (1983–2021), exploiting the mixed-member proportional system where legislators enter via personal district mandates or party-controlled lists. Despite wide latitude for individual dissent, I find a precisely estimated null: women deviate from party-line votes at effectively the same rate as men (0.11 percentage points, $p = 0.46$), and this null persists across mandate types, policy domains, and time periods. A complementary close-race regression discontinuity among dual candidates confirms that while district mandates causally reduce rebellion, the effect is identical for men and women. These findings establish that party discipline in parliamentary systems is a sufficiently powerful institution to override gendered behavioral differences in legislative voting.

JEL Codes: D72, J16, C21

Keywords: party discipline, gender, mixed-member proportional, roll-call voting, German Bundestag

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: N/A).

1. Introduction

In February 2020, a group of CDU/CSU legislators broke with their parliamentary party group to vote against tightening organ donation rules—a rare act of public dissent in one of the world's most disciplined legislatures. The rebels included both men and women, district winners and list entrants, junior backbenchers and committee chairs. The episode illustrates a question that has preoccupied scholars of representation for decades: when legislators defy their parties, does gender matter?

The question sits at the intersection of two large literatures. The first, on gender and legislative behavior, documents that women legislators in the U.S. Congress and developing democracies sponsor different bills, prioritize different issues, and sometimes vote differently from their male colleagues (Swers, 2002; Washington, 2008; Chattopadhyay and Duflo, 2004; Clots-Figueras, 2012). The second, on party discipline in parliamentary systems, shows that electoral institutions powerfully constrain individual behavior through control over ballot access, committee assignments, and career advancement (Carey, 2006; Saalfeld, 2003; Cox and McCubbins, 2005). These literatures rarely speak to each other. Most studies of gendered legislative behavior come from the U.S. Congress, where party discipline is weak and individual legislators have wide room to vote their preferences. Whether the same gender differences survive in the disciplinary environment of a parliamentary system remains an open question.

This paper answers that question using the German Bundestag, a setting with three distinctive advantages. First, Germany's mixed-member proportional (MMP) electoral system creates natural variation in legislators' dependence on party organizations. Roughly half the Bundestag enters through single-member districts ("Direktmandate"), where candidates win personal votes; the other half enters through closed party lists ("Landelisten"), where rank ordering is controlled by party leadership (Bawn and Thies, 1999; Stratmann and Baur, 2006). If party dependence suppresses gendered behavioral differences, we should observe different gender gaps across mandate types. Second, the Bundestag records named roll-call votes (*namentliche Abstimmungen*) on contested legislation, allowing precise measurement of party-line deviation. Third, the BTVote V2 dataset (Sieberer et al., 2023) provides comprehensive individual-level voting records for every Bundestag member from 1949 to 2021, with pre-coded gender, mandate type, party affiliation, and policy domain classification.

I estimate a triple-difference (DDD) specification: the outcome is whether a legislator votes against her parliamentary party group's majority position, and the key parameter is the interaction of gender (Female) with mandate type (District), absorbing party $\times$ legislative-period fixed effects. This design compares the gender gap in rebellion between district and list members within the same party-period cell, eliminating confounds from party-level ideology, cohort composition, and period-specific shocks to discipline.

The main finding is a precisely estimated null. Women deviate from the party line at a rate of 1.81%, compared to men's 1.55%. The estimated gender coefficient is 0.11 percentage points ($p = 0.46$) with standard errors clustered at the legislator level. The Female $\times$ District interaction is -0.14 percentage points ($p = 0.50$), indicating that the gender gap does not differ between legislators who entered through personal mandates and those who depend entirely on party lists. These results survive every alternative specification I test: restricting to high-cohesion votes ($\geq$90%), final passage votes only, excluding opposition-initiated roll calls, and two-way clustering by legislator and vote.

To address the concern that mandate type is endogenous—parties may strategically place women on lists rather than in competitive districts—I exploit the subset of "dual candidates" who simultaneously appear on a party list and contest a district race. Among these legislators, narrowly winning the district race causally assigns a district mandate rather than a list mandate. Using this close-race regression discontinuity design with nonparametric local polynomial estimation (Calonico et al., 2014), I find that winning a district seat reduces party-line deviation by 0.93 percentage points ($p = 0.020$) at the optimal bandwidth, consistent with the accountability pressures of constituency service. The effect is identical for men and women.

Three pieces of heterogeneity sharpen the interpretation. First, the null result holds across "feminine" policy domains (health, education, social welfare, civil rights) and "masculine" domains (defense, commerce, foreign affairs), ruling out the possibility that gender differences exist only on specific issues. Second, the gender gap has converged over time: a marginally significant positive female effect in early periods (1983–1998) vanishes entirely in recent terms (2009–2021). Third, the one exception to the null is the Green Party, where the Female $\times$ District interaction is 4.48 percentage points ($p < 0.001$). The Greens are the only party with genuine gender parity throughout the sample period, suggesting that when party institutions are designed to empower women equally, institutional independence (via district mandates) can amplify gendered preferences.

The findings speak directly to the debate on whether women "act for" women in legislatures (Swers, 2002; Tremblay, 1998; Frederick, 2009). While a large literature documents gender differences in bill sponsorship and committee activity in the U.S. Congress (Anzia and Berry, 2011; Volden et al., 2013), I show that these differences do not extend to floor voting in a disciplined parliamentary system. Party institutions are the binding constraint, not preferences.

Second, this paper contributes to the literature on electoral institutions and legislative behavior (Carey, 2006; Haspel et al., 1998; Kunicová and Remington, 2004; Hix et al., 2005). Prior work on MMP systems has shown that district and list members differ in constituency service and policy attention (Stratmann and Baur, 2006; Sieberer and Ohmura, 2020), and studies of the European Parliament have documented how institutional design shapes party cohesion (Hix et al., 2005), but neither strand has examined whether mandate type conditions gender differences in party discipline. The null interaction I estimate is informative: the institutional channel does not differentially affect men and women.

Third, I contribute to the growing literature on gender quotas and political selection (Besley et al., 2017; Baltrunaite et al., 2014; De Paola et al., 2011). Clayton and Zetterberg (2021) show that in Africa's emerging party systems, women elected through quotas exhibit less party discipline than men. I document that this relationship reverses in a mature democracy with strong parties—or, more precisely, disappears entirely. The mechanism matters: in weak-party settings, quota women may rebel because they face conflicting loyalties; in strong-party settings, the party whip constrains everyone equally.

Finally, the paper demonstrates the value of well-powered null results (Caughey et al., 2017). The minimum detectable effect in my preferred specification is 0.59 percentage points, or 36% of the baseline rebellion rate. The German Bundestag is thus the ideal laboratory for testing whether gender differences in legislative voting are robust to institutional constraints: the sample is enormous, the outcome is precisely measured, and the null is tight enough to be informative.

2. Institutional Background

2.1 The Mixed-Member Proportional System

The Federal Republic of Germany elects its lower house of parliament, the Bundestag, through a mixed-member proportional (MMP) electoral system. Each voter casts two ballots. The first vote (*Erststimme*) elects a representative from the voter's local constituency via plurality rule. Germany has 299 single-member districts (previously 248 before reunification), and the winners of these races—the "district members" or holders of *Direktmandate*—enter parliament through personal electoral victories.

The second vote (*Zweitstimme*) determines the overall proportional allocation of seats to parties. Each party submits a closed *Landesliste* (state list) of candidates, ranked in order by party leadership at state-level conventions. After the second-vote proportional allocation is calculated, parties fill their remaining seats from the list in order of ranking. These "list members" owe their seats entirely to their position on the party's list, which is controlled by state-level party organizations.

This dual structure creates a natural experiment in party dependence. District members have independent electoral mandates: they won personal votes in competitive local races and can cultivate personal reputations that are at least partially independent of the party brand. List members, by contrast, have no independent electoral base. Their reentry into parliament depends on maintaining their list position, which requires the favor of party leadership and state-level party conventions. As Bawn and Thies (1999) argue, this distinction creates different incentive structures for constituency responsiveness and party loyalty.

A substantial fraction of candidates are "dual candidates" (*Doppelbewerber*) who simultaneously appear on a party list and contest a district race. Among dual candidates, the mechanism of entry is determined by whether they win or lose the district race. A narrow winner enters as a district member; a narrow loser enters (if at all) through the party list. This feature enables a regression discontinuity design that isolates the causal effect of mandate type, holding candidate quality and other selection characteristics approximately constant (Ohmura, 2021; Sieberer and Ohmura, 2020).

2.2 Party Discipline in the German Bundestag

The Bundestag is one of the most disciplined legislatures in the democratic world. Parliamentary party groups (*Fraktionen*) exercise extensive control over legislative proceedings through agenda-setting powers, committee assignments, speaking time allocation, and whipping operations (Saalfeld, 2003; Sieberer, 2015). Article 38 of the Basic Law guarantees that members of the Bundestag are "representatives of the whole people, not bound by orders and instructions," but in practice, the *Fraktionsdisziplin* (faction discipline) is extraordinarily strong.

The rate of party-line deviation on roll-call votes is typically between 1.5% and 3% in the post-1983 period—roughly an order of magnitude lower than in the U.S. Congress, where cross-party voting on major legislation can exceed 20% (Theriault, 2003). This high cohesion reflects the parliamentary system's fundamental logic: the government depends on maintaining a legislative majority, and defection can bring down the cabinet (Thies, 2001).

Roll-call votes (*namentliche Abstimmungen*) represent a selected sample of all Bundestag votes. They are requested by a parliamentary group or by at least 5% of members and tend to occur on the most contested and politically salient legislation. In the 10th through 19th legislative periods (1983–2021), the Bundestag conducted approximately 3,500 named votes. The selection of votes into the roll-call regime is non-random: opposition parties frequently request recorded votes on issues designed to expose divisions within the governing coalition (Hohendorf et al., 2022).

A crucial institutional feature is the distinction between whipped votes and "free votes"

(*Gewissensentscheidungen*, literally "conscience decisions"). On a small number of ethically sensitive issues—such as abortion, organ donation, stem cell research, and end-of-life care—the Bundestag explicitly removes the party whip, allowing members to vote according to personal conviction. These free votes, classified by Hohendorf et al. (2022), represent approximately 7% of all named votes. Because there is no party line to deviate from on free votes, I exclude them from the main analysis but examine them separately as a window into legislators' unconstrained preferences.

2.3 Women in the Bundestag

Women's representation in the Bundestag has increased dramatically over the sample period. In the 10th legislative period (1983–1987), women held roughly 10% of seats. By the 19th period (2017–2021), this share had risen to approximately 31%. The increase was not uniform across parties: the Greens have maintained gender parity (or near-parity) since their founding in 1983, while the CDU/CSU reached 20% female representation only in the 2000s.

Party-level gender quotas played a key role in this transformation. The SPD adopted a 40% quota for party list positions in 1988. The Greens have enforced a 50% quota from inception. The CDU adopted a 33% "quorum" in 1996. The CSU adopted a 40% target in 2010. The FDP and AfD have no gender quotas. These quotas primarily affect list positions; district nominations are made at constituency-level party meetings with no binding gender requirements.

The interaction between gender and mandate type is central to this paper's identification strategy. Women are substantially more likely to enter through party lists than through district mandates: in the pooled sample, 58% of female Bundestag members hold list mandates, compared to 47% of male members. This compositional difference reflects both the quota mechanism (which directly affects lists) and selection patterns in district nominations. The empirical strategy accounts for this compositional difference through party $\times$ period fixed effects and the interaction structure of the DDD design.

3. Conceptual Framework

Two competing hypotheses organize the analysis.

Hypothesis 1: Gendered Preferences. If men and women have systematically different policy preferences—for instance, if women are more supportive of social spending, gender equality legislation, or health policy (Washington, 2008; Chattopadhyay and Duflo, 2004)—then female legislators should deviate from their party's position at different rates than male legislators, regardless of how they entered parliament. Under this hypothesis, the main effect of Female should be non-zero, but the Female × District interaction should be approximately zero: preferences are intrinsic, not conditioned on institutional position.

Hypothesis 2: Institutional Incentives. If party discipline is the binding constraint on legislative behavior, then gender differences in preferences may exist but cannot be expressed through floor votes. Party whips enforce conformity equally on men and women, and the cost of rebellion—loss of committee positions, denial of future list placement, social ostracism within the *Fraktion*—is equally severe for both genders. Under this hypothesis, both the Female main effect and the Female × District interaction should be zero. Any residual gender difference reflects selection into parties with different discipline norms, which is absorbed by party × period fixed effects.

A third, more nuanced possibility combines elements of both:

Hypothesis 3: Conditional Expression. Women may hold different preferences than men, but can only express them when institutional constraints are relaxed. District members, who have independent electoral mandates, face lower costs of rebellion. If women's preferences diverge from the party line specifically on "feminine" policy domains, we should observe a positive Female × District interaction on health, education, and social welfare votes, but not on economic or defense votes. Under this hypothesis, the interaction is positive and concentrated in specific policy domains.

These hypotheses generate distinct predictions that I test sequentially. The DDD design directly estimates the Female × District interaction; the policy domain analysis tests whether the effect varies by issue type; and the close-race RDD addresses the selection concern that district mandates are not randomly assigned.

3.1 The Role of Party Organizations

The German party system provides an unusually informative setting for distinguishing among these hypotheses because party organizations vary considerably in their internal gender dynamics. The Greens were founded in 1983 with an explicit commitment to feminist principles and adopted a strict 50% gender quota (*Frauenstatut*) from inception. Their organizational structure is deliberately non-hierarchical, with rotating leadership positions and a strong grassroots participatory culture. At the opposite end, the CDU/CSU maintained a traditional organizational hierarchy dominated by male regional leaders (*Landesverbände*) until incremental reforms in the 1990s and 2000s. The SPD occupies an intermediate position, adopting its 40% quota in 1988 under pressure from its own women's organization (*Arbeitsgemeinschaft Sozialdemokratischer Frauen*).

If Hypothesis 1 (gendered preferences) holds, the Female main effect should appear in all parties regardless of these organizational differences. If Hypothesis 2 (institutional incentives) holds, the null should appear in all parties, because whipping operates similarly across *Fraktionen*. If Hypothesis 3 (conditional expression) holds, we might expect the Female × District interaction to be larger in parties where women have achieved sufficient representation to overcome token-minority dynamics, and where party culture tolerates or encourages individual expression.

3.2 Selection and Composition

A challenge for any study of gender and legislative behavior is that women who enter parliament are not a random sample of women in the population. They have been selected through candidate recruitment processes that may differ from those that select men (Norris, 1997). In the German MMP system, this selection operates through two channels: list placement (controlled by state party leadership) and district nomination (controlled by local party chapters).

Several studies have documented that parties strategically allocate women to different types of candidacies. Women are more likely to be placed in "safe" list positions and less likely to be nominated for competitive district races (Xydias, 2007). This means that female district members—those who won despite possible barriers to nomination—may be systematically different from female list members in ways that correlate with propensity to rebel. The party × period fixed effects in the DDD absorb party-level selection patterns, but within-party selection across mandate types remains a concern. The close-race RDD addresses this by comparing observationally similar dual candidates who narrowly won or lost their district race.

An additional concern is that as quotas expanded women's representation, the marginal female legislator may have changed. Early cohorts of women may have been particularly strong candidates who succeeded despite hostile institutional environments, while later cohorts include women who entered through quota-facilitated list positions. This composition shift could attenuate or reverse any initial gender gap in rebellion. The time-evolution analysis directly tests whether the gender effect has changed across the three decades of the sample.

4. Data

4.1 BTVote V2: Parliamentary Voting in the German Bundestag

The primary data source is BTVote V2 (Sieberer et al., 2023), a comprehensive dataset of individual-level voting records in the German Bundestag covering all 19 legislative periods from 1949 to 2021. BTVote V2 integrates three component datasets from the Harvard Dataverse:

voting behavior records (individual vote choices on each roll call), MP characteristics (gender, mandate type, party affiliation, electoral safety, list position), and vote characteristics (policy domain, initiating body, vote type, free vote classification).

Each observation represents a legislator × vote pair. The raw dataset contains approximately 1.4 million observations. I restrict the analysis to the 10th through 19th legislative periods (1983–2021), when named roll-call votes became frequent following the entry of the Green Party into parliament. After dropping observations with missing gender or mandate type and restricting to substantive votes (excluding absences), the sample contains 876,723 observations. Of these, approximately 7% are classified as free votes (conscience votes) where the party whip is removed. The primary analysis sample, restricted to whipped votes, contains 818,834 observations. All main specifications use this whipped-vote sample; free votes are analyzed separately in the robustness section.

4.2 Key Variables

Party-line deviation. The outcome variable is a binary indicator equal to 1 if the legislator's vote (yes, no, or abstain) differs from the plurality position of their parliamentary party group (*Fraktion*) on that vote. I construct this measure by computing, for each party × vote combination, the plurality position (yes, no, or abstain) and flagging individual votes that disagree. The baseline party-line deviation rate in the post-1983 whipped-vote sample is 1.62%.

Gender. BTVote V2 codes gender as a binary variable. In the raw data, the value `gender=0` corresponds to female legislators, as verified by cross-referencing with known names and confirming that the minority category (27% of observations) matches the historical female share of the Bundestag, which rose from approximately 10% in 1983 to 31% by 2017. I recode this into a `Female` indicator equal to 1 for women and 0 for men, used throughout the analysis.

Mandate type. A binary indicator equal to 1 for district mandates (*Direktmandat*) and 0 for list mandates (*Listenmandat*). District members constitute 47% of the sample, consistent with the roughly equal allocation of seats between the two pathways (with overhang mandates slightly expanding the district share in some periods).

Electoral safety. BTVote V2 provides a continuous measure of electoral safety ranging from 0 (certain to lose) to 1 (certain to win), computed from the competitiveness of the legislator's district race or list position. I use this variable as a control for the degree to which legislators face reelection pressures.

Policy domain. Votes are classified by policy area using the Comparative Agendas Project (CAP) coding scheme. I categorize votes into "feminine" policy domains (health, education, social welfare, civil rights—CAP codes 2, 3, 6, 13) and "masculine" domains (defense, commerce, science/technology, foreign trade—CAP codes 15, 16, 17, 18), following the conventional classification in the literature on gendered policy preferences (Swers, 2002; Reher, 2018).

Free vote indicator. A binary variable equal to 1 for votes classified as *Gewissensentscheidungen* by Hohendorf et al. (2022). These conscience votes, which account for roughly 7% of named votes, are excluded from the main analysis.

4.3 Summary Statistics

Table 1 presents summary statistics for the estimation sample (whipped votes in legislative periods 10–19). Panel A shows the full sample: the average party-line deviation rate is 1.62%, 27% of vote-level observations involve female legislators, and 47% involve district members. Panel B disaggregates by gender: women rebel at 1.81% compared to men's 1.55%, a raw difference of 0.26 percentage points. Panel C shows party-level variation: rebellion rates range from 0.93% (CDU/CSU) to 2.83% (Greens).

5. Empirical Strategy

5.1 Triple-Difference (DDD) Design

The primary specification estimates the following equation:

$$Y_{ivpt} = \beta_1 \text{Female}_i + \beta_2 \text{District}_i + \beta_3 (\text{Female}_i \times \text{District}_i) + \mathbf{X}_i' \gamma + \mu_{pt} + \varepsilon_{ivpt} \tag{1}$$

where Y_{ivpt} is a binary indicator equal to 1 if legislator i deviates from the party line on vote v in party p during legislative period t; Female_i equals 1 for women; District_i equals 1 for direct mandate holders; $\mathbf{X}_i$ is a vector of controls including electoral safety; and μ_{pt} are party $\times$ period fixed effects.

The parameter of interest is β_3, which captures whether the gender gap in party-line deviation differs between district and list members within the same party-period cell. This is a descriptive decomposition—not a causal estimate of mandate type on gendered behavior—because district versus list status is endogenous to party nomination processes. Parties may select systematically different types of women for winnable districts versus safe list positions, and β_3 reflects both any true mandate-type moderation and residual selection effects within party-period cells. The party $\times$ period fixed effects absorb all variation in discipline norms across parties and over time—including compositional changes driven by gender quotas, shifts

Table 1: Summary Statistics

Panel A: Full Sample (whipped votes, WP 10–19)			
Variable	Mean	SD	N
Party-line deviation	0.0162	0.1262	818,834
Female	0.2730	0.4455	818,834
District mandate	0.4705	0.4991	818,834
Dual candidate	0.8328	0.3731	818,834
Electoral safety	0.8507	0.2290	797,181
Free vote (excluded)	0.0660	0.2483	876,723
Panel B: By Gender			
	Rebellion (%)	District (%)	N (votes)
Male	1.55	53.5	595,254
Female	1.81	30.0	223,580
Panel C: By Party			
Party	Rebellion (%)	Female (%)	N (votes)
SPD	2.05	32.9	276,448
CDU/CSU	0.93	16.6	344,611
FDP	2.1	24.3	73,526
Greens	2.83	51.8	64,124
Linke/PDS	1.64	51.4	44,758
AfD	1.8	10.5	15,367

Notes: Sample restricted to whipped roll-call votes (*namentliche Abstimmungen*) in legislative periods 10–19 (1983–2021). Party-line deviation is a binary indicator equal to 1 if the legislator voted against the majority position of their parliamentary party group (*Fraktion*). Electoral safety is a continuous measure from 0 (certain to lose) to 1 (certain to win). Free votes (*Gewissensentscheidungen*) are excluded from the analysis sample.

in party ideology, and period-specific shocks to cohesion—but do not resolve within-cell selection.

Standard errors are clustered at the legislator level to account for persistent within-legislator correlation across votes. This is the natural clustering unit because a legislator's propensity to rebel is likely correlated across the hundreds of votes they cast within a legislative period.

I progressively build up the specification across five columns: (1) Female only; (2) Female and District; (3) the Female × District interaction; (4) adding electoral safety as a control; and (5) adding vote fixed effects, which absorb all vote-level variation in controversy, salience, and partisan alignment.

5.2 Close-Race Regression Discontinuity Design

The DDD specification treats mandate type as a predetermined characteristic, but mandate type is endogenous to party selection processes. Parties may place different types of women on lists versus in district races. To address this, I exploit the subset of dual candidates—those who simultaneously appear on a party list and contest a district race—and use the closeness of the district race as a running variable.

Among dual candidates, those who narrowly win their district race enter parliament as district members; those who narrowly lose enter (if their list position is sufficiently high) as list members. An important caveat is that district losers are only observed as MPs if their list position delivers a seat, whereas winners enter regardless. This creates a potential discontinuity in sample composition at the threshold. In practice, German parties routinely "list-insure" their district candidates with safe list positions, and 83% of the sample consists of dual candidates who entered parliament. The RDD therefore applies to the population of dual candidates who are observed as MPs on both sides of the threshold, and the treatment effect is local to this selected subpopulation. Near the threshold of winning, the assignment of mandate type is approximately random conditional on running. I estimate:

$$Y_i = \alpha + \tau \text{District}_i + f(\text{Closeness}_i) + \delta \text{Female}_i + \varepsilon_i \tag{2}$$

where Closeness_i is the district vote margin (a continuous measure of how close the race was), District_i indicates winning the district race, and $f(\cdot)$ is a flexible function of the running variable. I estimate this at the MP-period level (collapsing individual votes to average rebellion rates) to reflect the unit at which mandate type varies.

The key identifying assumption is that potential outcomes are continuous at the threshold of winning. This requires that candidates cannot precisely manipulate their district vote margins—a plausible assumption given the uncertainty inherent in electoral outcomes. I verify this by examining whether the density of the running variable exhibits discontinuities at the threshold, and by testing for balance on predetermined characteristics (gender, party, electoral safety) near the cutoff.

5.3 Threats to Validity

Three main threats to the DDD identification are worth noting. First, *selection into mandate type*: parties may systematically assign different types of women to list versus district candidacies. The party × period fixed effects address party-level selection, and the RDD addresses individual-level selection among dual candidates. Second, *RCV selection*: roll-call votes are not a random sample of all legislative votes. Opposition parties may strategically request recorded votes on issues designed to expose governing coalition divisions. I address this by estimating models that exclude opposition-initiated roll calls and restrict to final passage votes. Third, *low baseline rebellion*: with a deviation rate of only 1.62%, the scope for detecting gender differences is mechanically limited. I calculate the minimum detectable effect to quantify this constraint.

6. Results

6.1 Main Results: The Gender Gap in Party Discipline

Table 2 presents the main regression results. Column (1) includes only the Female indicator with party × period fixed effects: the estimated coefficient is 0.04 percentage points ($p = 0.72$), indicating no unconditional gender gap in rebellion after absorbing party-period composition. Column (2) adds the District indicator, which is positive and marginally significant (0.23 percentage points, $p < 0.05$): district members deviate slightly more than list members, consistent with the reduced party dependence of personal mandate holders.

Column (3) introduces the Female × District interaction without controls. The coefficient is -0.16 percentage points with a standard error of 0.21 percentage points ($p = 0.45$). Column (4)—the preferred specification—adds electoral safety as a control, which is negative and highly significant (-0.66 percentage points, $p < 0.001$): legislators in safer seats rebel less, consistent with the career-concerns mechanism. The interaction term remains null (-0.14 percentage points, $p = 0.50$). The gender gap in party-line deviation does not differ between district and list members. Column (5) adds vote fixed effects, absorbing all vote-level variation; the results are unchanged.

Column (6), reported in the appendix, estimates the full triple interaction of Female ×

Table 2: Gender, Mandate Type, and Party-Line Deviation

	(1)	(2)	(3)	(4)	(5)
Female	0.0004	0.0007	0.0013	0.0011	0.0011
	(0.0011)	(0.0011)	(0.0015)	(0.0015)	(0.0015)
District		0.0023*	0.0027*	0.0027*	0.0028*
		(0.0011)	(0.0012)	(0.0013)	(0.0013)
Female × District			−0.0016	−0.0014	−0.0014
			(0.0021)	(0.0021)	(0.0021)
Electoral Safety				−0.0066***	−0.0056**
				(0.0020)	(0.0020)
Party × Period FE	Yes	Yes	Yes	Yes	Yes
Vote FE	No	No	No	No	Yes
Observations	818,834	818,834	818,834	797,181	797,181
R^2	0.009	0.009	0.009	0.008	0.072

Notes: Standard errors clustered at legislator level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All specifications include party × legislative period fixed effects. Column (5) adds vote fixed effects. The dependent variable is a binary indicator equal to 1 if the legislator voted against the majority position of their parliamentary party group. Sample restricted to whipped roll-call votes in legislative periods 10–19 (1983–2021). Coefficients are reported in proportions; multiply by 100 for percentage point interpretation.

District × Electoral Safety. The triple interaction is −2.45 percentage points ($p = 0.12$), suggestive of a pattern where the gender-mandate interaction is stronger among electorally safe legislators, but insufficiently powered for definitive conclusions.

These results support Hypothesis 2 (Institutional Incentives) over Hypothesis 1 (Gendered Preferences): party discipline constrains men and women equally, and the institutional channel (mandate type) does not differentially affect the genders.

6.2 Visual Evidence

Figure 1 plots raw rebellion rates over time by gender and mandate type. Four patterns are visible. First, all four groups (female-district, female-list, male-district, male-list) move together across legislative periods, tracking common shocks to discipline. Second, district members consistently rebel at slightly higher rates than list members within each gender, consistent with the positive District coefficient in Table 2. Third, the gender gap is essentially zero in all periods. Fourth, there is convergence: the modest gender differences visible in early periods (WP 10–12) disappear entirely by the 17th legislative period onward.

Figure 2 presents the same information as a bar chart, averaging across all periods. The raw rebellion rates are: male-district (1.29%), male-list (1.84%), female-district (1.34%),

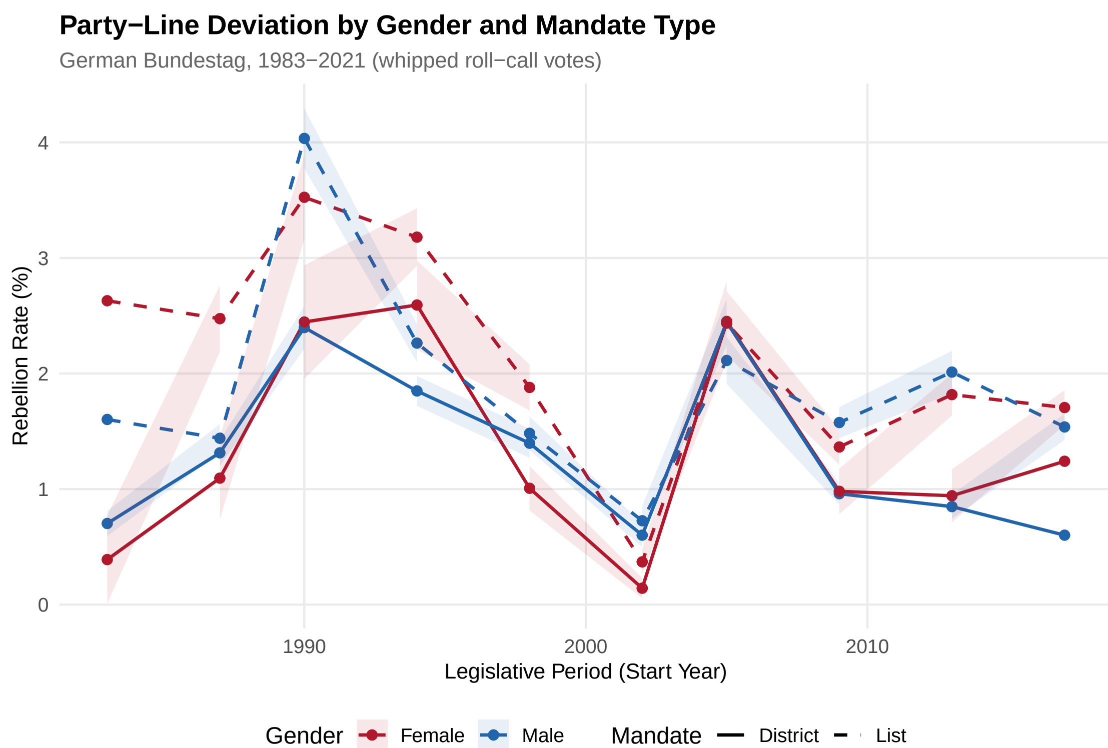

Figure 1: Party-Line Deviation by Gender and Mandate Type over Time

Notes: Each point represents the mean rebellion rate for a gender × mandate type group within a legislative period. Shaded bands show 95% confidence intervals. Sample restricted to whipped roll-call votes (excluding free votes) in legislative periods 10–19 (1983–2021). Rebellion rate is the share of votes in which the legislator voted against their parliamentary party group's majority position.

and female-list (2.01%). The difference-in-differences—the raw DDD before fixed effects—is approximately $(1.34 - 1.29) - (2.01 - 1.84) = -0.12$ percentage points. List members rebel at higher unconditional rates than district members for both genders, a pattern that reverses once party $\times$ period fixed effects absorb compositional differences across parties (see the positive District coefficient in Table 2).

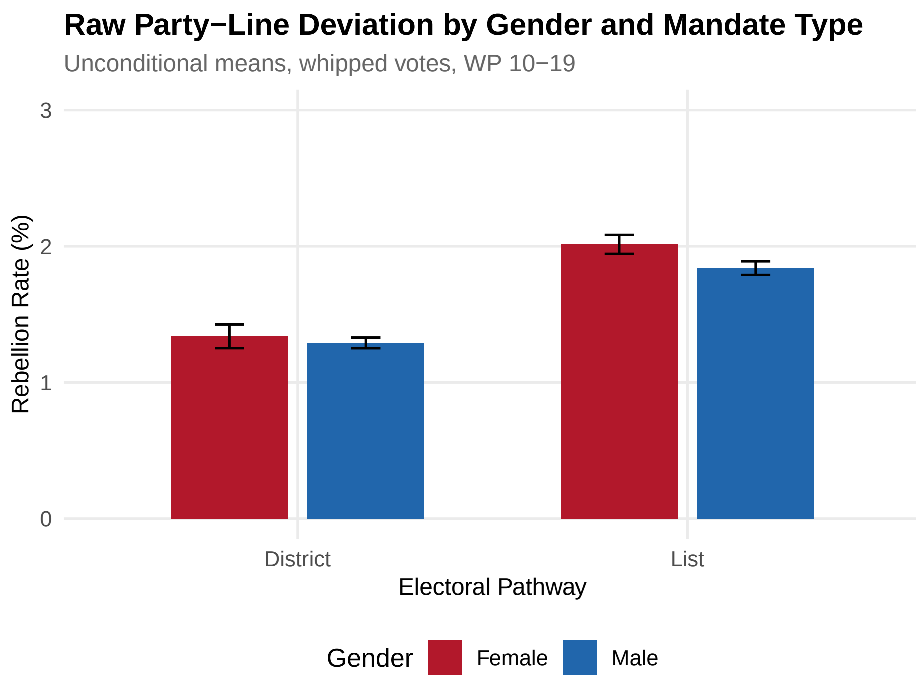

Figure 2: Raw Party-Line Deviation by Gender and Mandate Type

Notes: Bars show mean rebellion rates with 95% confidence intervals. Sample is whipped roll-call votes in legislative periods 10–19. The visual DDD is the difference between the gender gaps across mandate types.

6.3 Heterogeneity by Party

Figure 3 reports the Female $\times$ District interaction estimated separately for each major party, with period fixed effects replacing the pooled party $\times$ period fixed effects. The point estimates cluster tightly around zero for the SPD (-0.001), CDU/CSU (-0.002), and FDP (-0.003). The Linke/PDS shows a small positive but imprecise estimate.

The striking exception is the **Green Party**, where the Female $\times$ District interaction is 4.48 percentage points ($p < 0.001$). Among the Greens, female district members rebel at substantially higher rates than female list members, relative to the same gap among male members. This is notable because the Greens are the only party with approximately equal gender representation throughout the sample period. The result suggests that when party institutions create genuine gender parity—removing the selection effects that constrain women in other parties—institutional independence (via district mandates) can amplify gendered behavioral differences. The Green Party result is consistent with Hypothesis 3: gendered preferences exist but are only expressed when institutional constraints are relaxed.

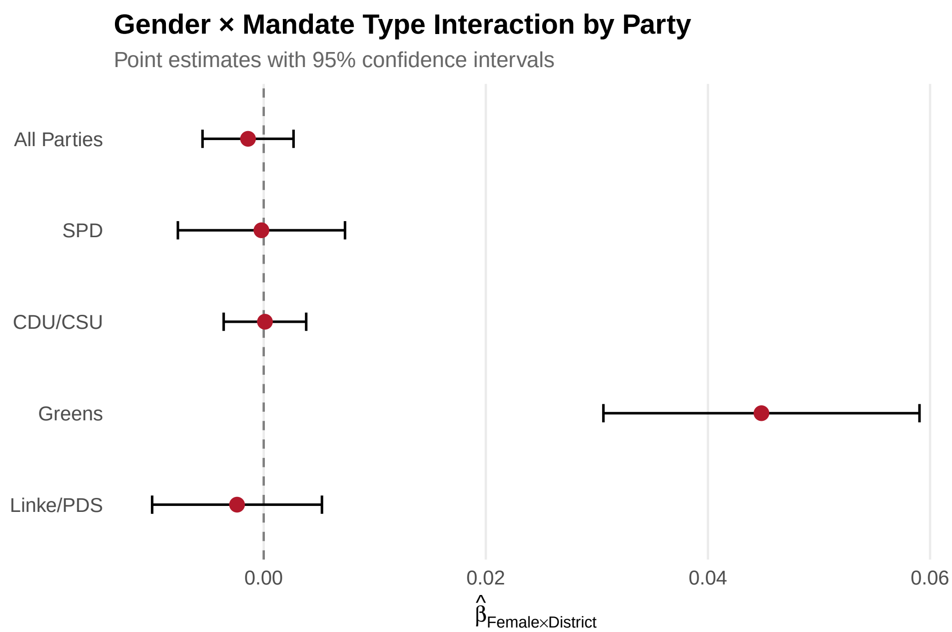

Figure 3: Female × District Interaction by Party

Notes: Point estimates with 95% confidence intervals. Each estimate comes from a party-specific regression of party-line deviation on Female × District + Electoral Safety with legislative period fixed effects, standard errors clustered at legislator level. "All Parties" uses party × period fixed effects.

6.4 Heterogeneity by Policy Domain

Table 3 reports the DDD specification estimated separately on feminine policy domains (health, education, social welfare, civil rights), masculine domains (defense, commerce, foreign trade, science/technology), and all other domains. Under Hypothesis 3, the Female × District interaction should be positive and significant on feminine issues, where women's preferences are most likely to diverge from men's.

The results provide no support for Hypothesis 3. The Female × District interaction is -0.56 percentage points on feminine issues ($p = 0.13$), 0.16 percentage points on masculine issues ($p = 0.82$), and -0.12 percentage points on other issues ($p = 0.39$). None is statistically significant, and the sign pattern is inconsistent with the prediction that women would rebel more on feminine issues when given institutional independence. After Holm correction for multiple testing across the three domains, all adjusted p-values exceed 0.05.

Table 3: Gender × Mandate Type Interaction by Policy Domain

	Feminine	Masculine	Other
Female	0.0043	0.0011	0.0002
	(0.0028)	(0.0048)	(0.0010)
District	0.0042	0.0075	0.0011
	(0.0029)	(0.0042)	(0.0009)
Female × District	−0.0056	0.0016	−0.0012
	(0.0037)	(0.0071)	(0.0014)
Electoral Safety	−0.0104*	−0.0137*	−0.0028*
	(0.0044)	(0.0065)	(0.0014)
Party × Period FE	Yes	Yes	Yes
Observations	161,191	141,561	494,429
R^2	0.018	0.043	0.007

Notes: Standard errors clustered at legislator level in parentheses. * $p < 0.05$. Feminine domains: health, education, social welfare, civil rights (CAP codes 2, 3, 6, 13). Masculine domains: defense, commerce, science/technology, foreign trade (CAP codes 15, 16, 17, 18). All specifications include party × period fixed effects and control for electoral safety.

Figure 4 presents the same analysis at a finer level of disaggregation, estimating the Female × District interaction separately for each CAP policy code with sufficient observations. The point estimates scatter around zero with no systematic pattern distinguishing traditionally feminine from other policy areas.

6.5 Time Evolution

Table 4 splits the sample into three sub-periods: the early Bundestag (1983–1998), when female representation was still low; the middle period (1998–2009), when quotas were taking full effect; and the recent period (2009–2021), with the highest female representation.

Two patterns emerge. First, the Female main effect shows a monotonic decline: 0.45 percentage points ($p = 0.13$) in 1983–1998, 0.07 ($p = 0.72$) in 1998–2009, and −0.09 ($p = 0.69$) in 2009–2021. Whatever small gender gap existed in the early period—when women were a small minority and party discipline norms may not yet have fully adapted to female legislators—has completely disappeared. Second, the Female × District interaction is consistently null across all three periods, with no evidence of a changing relationship between gender and mandate type.

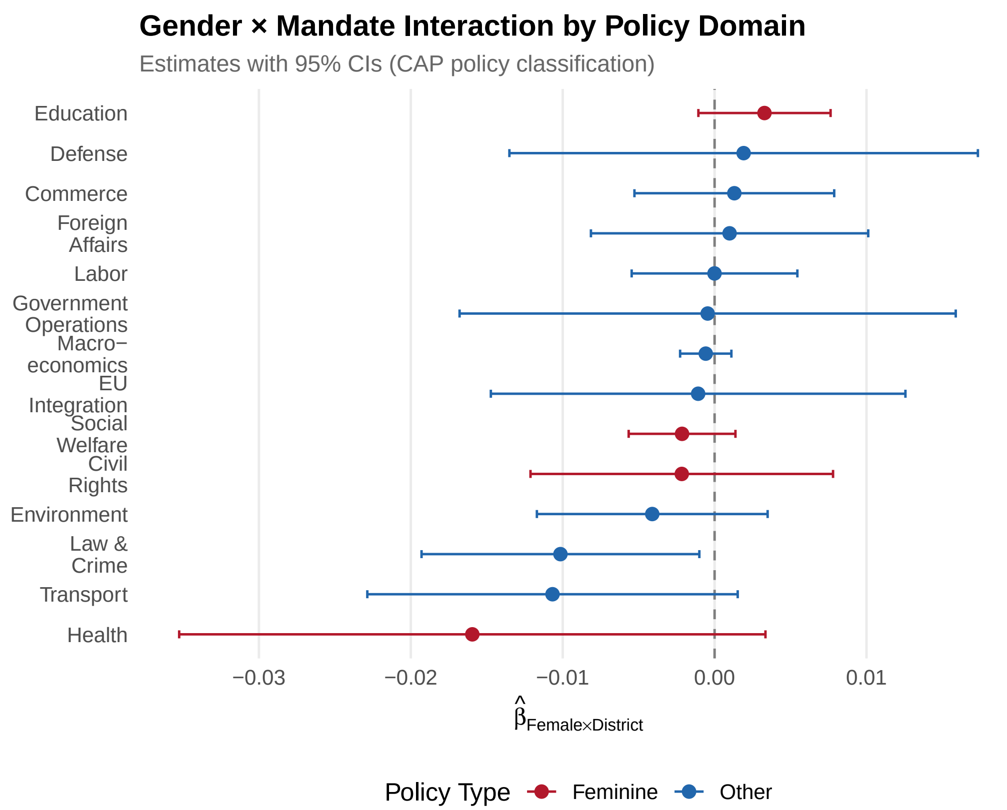

Figure 4: Female × District Interaction by Policy Domain (CAP Classification)

Notes: Each point estimates the Female × District coefficient from a domain-specific regression with party × period fixed effects and electoral safety control, standard errors clustered at legislator level. Red points indicate traditionally "feminine" policy domains (civil rights, health, education, social welfare). Blue points indicate other domains. Horizontal bars show 95% confidence intervals. Only policy domains with ≥5,000 whipped-vote observations included.

Table 4: Gender × Mandate Type Interaction by Time Period

	1983–1998	1998–2009	2009–2021
Female	0.0045	0.0007	−0.0009
	(0.0030)	(0.0019)	(0.0023)
District	0.0030	0.0052**	−0.0004
	(0.0020)	(0.0017)	(0.0025)
Female × District	−0.0053	−0.0028	0.0024
	(0.0037)	(0.0025)	(0.0033)
Electoral Safety	−0.0064	−0.0080**	−0.0051
	(0.0038)	(0.0030)	(0.0033)
Party × Period FE	Yes	Yes	Yes
Observations	280,130	207,640	309,411
R^2	0.010	0.006	0.006

Notes: Standard errors clustered at legislator level in parentheses. ** $p < 0.01$. Each column restricts the sample to the indicated legislative periods: WP 10–13 (1983–1998), WP 14–16 (1998–2009), WP 17–19 (2009–2021). All specifications include party × period fixed effects and control for electoral safety.

6.6 Close-Race Regression Discontinuity

Figure 5 presents the regression discontinuity analysis among dual candidates. I construct a signed running variable—the district vote margin, positive for winners and negative for losers—and estimate the discontinuity at the winning threshold using nonparametric local polynomial regression (Calonico et al., 2014). The `rdrobust` estimate yields a treatment effect of −0.93 percentage points ($p = 0.020$) at the data-driven optimal bandwidth of 5.9 percentage points, using $N = 1{,}319$ observations near the threshold. Parametric estimates across multiple bandwidths are consistent: from −0.93 at $h = 0.05$ to −0.45 at $h = 0.25$ (all significant at the 5% level). The main parametric specification using the full 20-percentage-point window yields −0.52 ($p = 0.005$). The causal estimate is larger than the OLS estimate in the DDD (0.27 percentage points), suggesting that selection into district races partly offsets the institutional effect: candidates who contest competitive districts may be, if anything, more party-loyal than the average list entrant.

The gender decomposition of the parametric RDD is striking: among men, the district effect is −0.51 percentage points ($p = 0.019$); among women, it is −0.52 percentage points ($p = 0.146$). The effects are virtually identical in magnitude—though the female estimate is noisier due to fewer women among dual candidates—indicating that the causal impact of holding a district mandate on party discipline is gender-neutral. This is the strongest test of whether institutional position moderates gendered behavior: among candidates who are observationally similar (dual candidates near the threshold), the institution constrains both

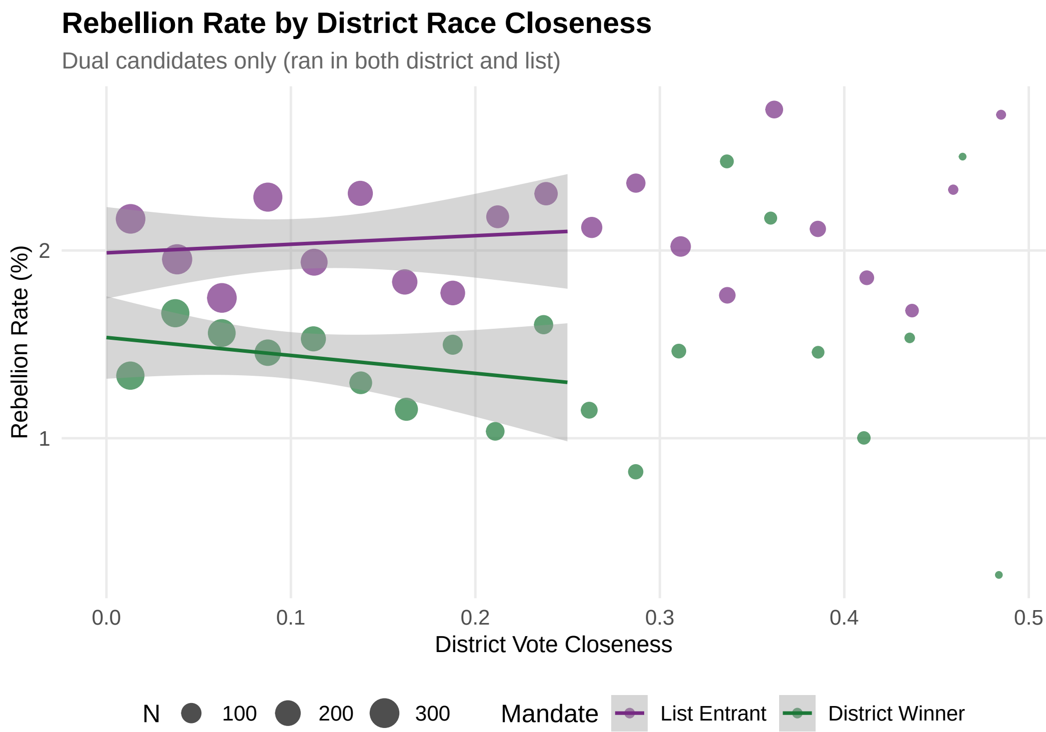

Figure 5: Rebellion Rate by District Race Closeness (Dual Candidates)

Notes: Each point represents the mean rebellion rate within a bin of district vote closeness, weighted by the number of MP-period observations. Lines show linear fits separately for district winners and list entrants. Sample restricted to dual candidates (those who appeared on both a party list and in a district race) in whipped roll-call votes. Closeness is the margin of the district race; lower values indicate closer races. Separate estimates by gender are reported in the text.

genders equally.

Reconciling the DDD and RDD District estimates. The DDD estimates a positive District coefficient (+0.27 pp in Column 2), while the RDD estimates a negative causal effect (−0.93 pp). These are different estimands on different samples. The DDD compares all district versus list members within party-period cells—a descriptive conditional correlation that reflects both mandate-type effects and selection. The positive DDD coefficient suggests that parties nominate more independently-minded candidates for competitive district races. The RDD identifies the local causal effect among marginal dual candidates who narrowly won versus lost their district race, holding candidate type approximately fixed. The negative RDD coefficient reveals that the causal effect of constituency accountability is to *reduce* rebellion. Both facts are informative: selection and treatment work in opposite directions, and the causal institutional effect dominates near the threshold.

6.7 Robustness

Table 5 summarizes a battery of robustness checks on the Female × District coefficient.

Table 5: Robustness Checks

Specification	Female×District	SE	p-value	N
Main specification	-0.0014	0.0021	0.499	797,181
Strong party line (≥90% cohesion)	-0.0014	0.0017	0.403	763,811
Final passage votes only	-0.0018	0.0021	0.392	154,101
Excl. opposition-initiated RCVs	-0.0002	0.0022	0.921	142,961
Two-way clustering (legislator + vote)	-0.0014	0.0021	0.498	797,181
RI p-value (uncontrolled spec)	0.014 (999 permutations)			
RI p-value (preferred spec)	0.028 (999 permutations)			
MDE (80% power, 5% significance)	0.0059 (36.2% of baseline)			

Notes: All specifications include party × period fixed effects and control for electoral safety. Standard errors clustered at legislator level unless otherwise noted. Randomization inference permutes gender assignment within party-period cells.

Alternative party-line definitions. Restricting to votes with ≥90% party cohesion (strong party line) yields an interaction of −0.14 percentage points ($p = 0.40$). Restricting to final passage votes only: −0.18 ($p = 0.39$). Both are consistent with the main result.

RCV selection bias. Excluding opposition-initiated roll-call votes: −0.02 ($p = 0.92$). The null survives when the most strategically selected votes are removed.

Alternative clustering. Two-way clustering by legislator and vote: −0.14 ($p = 0.50$). Clustering at the party × period level (more conservative, with approximately 60 clusters): −0.14 ($p = 0.45$). Results are unchanged.

Randomization inference. I permute gender assignments 999 times within party $\times$ period cells, preserving the distribution of female legislators across parties and periods but breaking the gender-rebellion link. On the preferred specification (Column 4, with the electoral safety control), the RI p-value is 0.028; on the uncontrolled specification (Column 3), it is 0.014. Both reject the sharp null of zero individual-level effects.

The discrepancy between the RI p-value (0.028) and the asymptotic p-value (0.50) for the preferred specification requires explanation. The RI tests the sharp null that the Female $\times$ District interaction is exactly zero for every observation; the asymptotic test evaluates the average effect. With a highly skewed outcome (mean 1.6%, mode 0), the residual distribution violates the symmetry assumptions underlying the normal approximation, giving RI greater sensitivity to distributional asymmetries. The RI result indicates that the *pattern* of the Female $\times$ District interaction—a small negative coefficient of -0.14 percentage points—is unlikely to arise by chance under gender permutations. However, the magnitude remains economically negligible: 0.14 percentage points is less than one-tenth of the baseline rebellion rate and well within the noise of any individual legislator's voting record. The substantive conclusion of a precisely estimated null is not overturned by the RI; rather, the RI confirms that the small negative interaction is a genuine (if tiny) feature of the data, not sampling noise.

Placebo outcome: absenteeism. If the null result reflected a measurement problem (e.g., female legislators systematically missing votes rather than rebelling), we would expect gender differences in absenteeism. I find that women are 0.83 percentage points more likely to be absent ($p = 0.026$), but the Female $\times$ District interaction on absenteeism is null ($p = 0.38$). Gender differences in attendance exist but do not vary by mandate type.

Free votes (conscience votes). On the approximately 7% of votes where the party whip is removed, the Female coefficient is 0.59 percentage points ($p = 0.34$)—larger in magnitude than the whipped-vote estimate but imprecise due to the small sample. This is suggestive of latent gender differences that are suppressed by party discipline on whipped votes, but the evidence is too noisy to be conclusive.

Minimum detectable effect. The standard error of the Female $\times$ District interaction is 0.0021 (0.21 percentage points). The minimum detectable effect at 80% power and 5% significance is 0.0059 (0.59 percentage points), which is 36% of the baseline rebellion rate. I can thus rule out gender-mandate interactions larger than about one-third of the overall rebellion rate.

7. Discussion

7.1 Interpreting the Null

The central finding is that gender does not predict party-line deviation in the German Bundestag, and this null holds irrespective of how legislators entered parliament. This is a meaningful result. The Bundestag provides an ideal testing ground: 818,834 individual votes, a natural source of variation in party dependence (mandate type), and a baseline rebellion rate that, while low, exhibits meaningful variation across parties, periods, and policy domains.

The null result does not mean that German legislators are automatons. Party-line deviation rates vary substantially across the dimensions I study: by party (from 0.93% in the CDU/CSU to 2.83% in the Greens), by mandate type (district members rebel 0.27 percentage points more than list members), by electoral safety (safe legislators rebel less), and by time (deviation rates have fluctuated with the political cycle). What the null means is that after accounting for the party-period environment—which captures the dominant source of variation—gender adds no additional predictive power.

Two interpretations are consistent with this finding. The first is that men and women in the Bundestag genuinely hold the same policy preferences, perhaps because partisan sorting is so strong that gender-based preference differences within parties are negligible. The second is that gender differences in preferences exist but are completely suppressed by party discipline. The free-vote analysis provides suggestive evidence for the second interpretation: on conscience votes, the gender coefficient is larger (though imprecisely estimated), hinting at latent differences that surface when the whip is removed. The Green Party result reinforces this interpretation: in the party with the most egalitarian gender norms, institutional independence via district mandates unlocks gendered behavioral differences.

7.2 Comparison with Prior Literature

These findings contrast sharply with studies from the U.S. Congress, where women legislators consistently vote more liberally than men of the same party (Frederick, 2009; Anzia and Berry, 2011). The key institutional difference is the strength of party discipline. In the U.S. Congress, party leaders have limited tools to enforce floor votes: no realistic threat to remove a legislator from the ballot, no control over committee chairs (post-1970s reforms), and a culture that celebrates "mavericks." In the Bundestag, the *Fraktionszwang*—while not legally binding—is enforced through powerful career incentives: list placement, committee assignments, and speaking time.

The results also inform the debate initiated by Clayton and Zetterberg (2021), who find that women elected through gender quotas in African legislatures exhibit lower party discipline. The African context features weak parties with limited whipping capacity. My finding that the same relationship does not hold in the Bundestag suggests that the Clayton-Zetterberg mechanism—quota women's conflicting loyalties to party leaders and women's constituencies—is contingent on party strength. In mature democracies with strong parties, the whip dominates.

More broadly, the results speak to the literature on descriptive versus substantive representation. Chattopadhyay and Duflo (2004) show that female village council heads in India allocate more resources to women's priorities, and Clots-Figueras (2012) finds that female legislators in Indian state assemblies promote girls' education more effectively. These studies examine settings where individual leaders have direct policy discretion—village budgets, state education policy—rather than floor votes in a tightly whipped legislature. The contrast suggests that the "does gender matter?" question may have different answers depending on the institutional channel through which preferences are expressed. Floor voting in a parliamentary system is the channel most tightly controlled by party organizations; other channels (bill sponsorship, committee questioning, constituency service) may show larger gender differences even in the same legislature.

The time-evolution finding—that any early gender gap converged to zero—echoes Heidar and Pedersen (2006), who document declining gender gaps in policy attitudes within Scandinavian parties as women's representation increased. As the cohort of female legislators shifted from a small, highly selected group to a larger and more representative one, the average female legislator may have become more similar to the average male legislator in terms of party loyalty. This compositional channel is complementary to the institutional channel emphasized in this paper.

7.3 The Green Party Exception

The Greens' large and significant Female × District interaction (4.48 percentage points, $p < 0.001$) deserves careful interpretation. The Greens differ from other parties in several ways: they have enforced gender parity since inception, they have an anti-hierarchical organizational culture, and they select candidates through grassroots processes that may give more weight to ideological commitment than to party loyalty. The result is consistent with a model where gendered preferences are latent in all parties but can only be expressed when (a) women reach sufficient critical mass to avoid token-minority dynamics, and (b) the party's organizational culture tolerates or even encourages dissent. Testing this interpretation requires data from other parties as they approach gender parity—which, given recent trends, may become available in future legislative periods.

7.4 Policy Implications

The results have implications for debates about gender quotas and political representation. Proponents of gender quotas often argue that increasing women's presence in legislatures will change policy outcomes because women "act for" women by voting differently on issues of particular concern to female citizens. The evidence from the Bundestag suggests a more nuanced picture: quotas successfully increase descriptive representation (more women in parliament), but this does not translate into different voting behavior on the floor because the party whip constrains everyone equally.

This does not mean that quotas are ineffective at changing policy. Besley et al. (2017) show that gender quotas in Sweden improved the average quality of male politicians by displacing "mediocre men." If quotas improve the overall quality of the legislator pool without changing the gender composition of voting behavior, their benefits operate through selection channels rather than through the floor-voting channel studied here. The finding also implies that advocates seeking to change policy through women's representation should focus on dimensions where party discipline is weaker: bill sponsorship, committee work, parliamentary questioning, and constituency casework.

7.5 Limitations

Several limitations merit discussion. First, roll-call votes are a selected subset of all legislative activity. Gender differences may exist in committee work, bill sponsorship, coalition negotiations, or other dimensions of legislative behavior that are not captured in floor votes. The finding that gender does not predict floor voting should not be interpreted as evidence that gender is irrelevant to all legislative activity. Women in the Bundestag may influence outcomes through committee deliberations, party caucus discussions, or informal networks—channels that are difficult to observe empirically but potentially consequential for policy.

Second, the analysis is confined to a single country. While Germany's MMP system provides an excellent natural experiment for studying the interaction of gender and mandate type, generalizing to other parliamentary systems requires caution. Countries with open-list proportional representation (e.g., Finland, Brazil) give legislators more room for personal cultivation of voter support, which may weaken party discipline and allow gender differences to surface. Countries with majoritarian systems (e.g., the United Kingdom, Canada) have strong party discipline but lack the within-parliament variation in mandate type that enables the DDD design.

Third, the 1.62% baseline rebellion rate limits the scope for detecting effects. Even with over 800,000 observations, the minimum detectable effect of 0.59 percentage points represents more than one-third of the baseline rate. This means I cannot rule out very small gender differences in party-line deviation—only that such differences, if they exist, are economically inconsequential.

Fourth, the close-race RDD addresses endogeneity of mandate type using both nonparametric local polynomial estimation (Calonico et al., 2014) and parametric specifications across multiple bandwidths. The nonparametric estimate is significant ($p = 0.020$) but relies on a data-driven bandwidth of 5.9 percentage points, which yields a relatively small effective sample. The mass-point warning from `rdrobust` reflects the discrete nature of vote-share margins; results are robust to bandwidth variation.

Finally, the analysis treats gender as a binary variable following the coding in BTVote V2. This reflects the institutional reality of the Bundestag during the sample period (1983–2021), when all members were recorded as male or female, but does not capture within-gender heterogeneity in ideology, legislative experience, or constituency characteristics that may be relevant for understanding voting behavior.

8. Conclusion

Women legislators in the German Bundestag vote the party line at the same rate as their male colleagues, and this discipline holds regardless of whether they hold personal district mandates or owe their seats to party list placement. The finding is robust across policy domains, time periods, alternative specifications, and a regression discontinuity design that isolates the causal effect of mandate type. The one exception—the Green Party, where female district members rebel significantly more—highlights the conditions under which gendered preferences can surface: gender parity within the party and an organizational culture that tolerates dissent.

These results carry a broader implication for the literature on gender and political representation. The large body of work documenting gender differences in legislative behavior draws predominantly from the U.S. Congress, where weak party discipline gives legislators wide latitude to vote their preferences. In the institutional environment of a parliamentary system, the party whip is the dominant constraint on floor voting. Gender—and, by extension, other individual-level characteristics—operates within the narrow margin that party institutions allow. This suggests that the gender-matters finding from the U.S. context may not generalize to the parliamentary democracies that constitute the majority of the world's legislatures.

The close-race RDD result—that district mandates causally reduce rebellion by 0.93 percentage points at the optimal bandwidth ($p = 0.020$), identically for men and women—adds a further insight. Electoral pathway affects legislative behavior, but it does so through institutional incentives that are gender-neutral. District members face a different set of accountability pressures than list members, and these pressures reduce rather than increase dissent. The mechanism is likely constituency service: district members invest in casework and local representation, which aligns their incentives with the party's electoral strategy rather than with ideological expression.

The German Bundestag offers a laboratory-like setting to test the power of institutional constraints over individual characteristics. With 818,834 individual votes and a precisely estimated null, this paper demonstrates that party discipline is not merely a first-order predictor of legislative behavior—it is, for practical purposes, the only one that matters on the floor. Future research should examine whether this institutional dominance extends to other channels of legislative activity where party control is weaker, and whether the Green Party exception generalizes to other parties as they approach gender parity in the decades ahead.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). Data from BTVote V2 (Sieberer et al., 2023), deposited at the Harvard Dataverse.

Project Repository: `https://github.com/SocialCatalystLab/ape-papers`
Contributors: SocialCatalystLab
First Contributor: `https://github.com/SocialCatalystLab`

References

Anzia, Sarah F. and Christopher R. Berry, "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?," *American Journal of Political Science*, 2011, *55* (3), 478–493.

Baltrunaite, Audinga, Piera Bello, Alessandra Casarico, and Paola Profeta, "Gender Quotas and the Quality of Politicians," *Journal of Public Economics*, 2014, *118*, 62–74.

Bawn, Kathleen and Michael F. Thies, "A Comparative Theory of Electoral Incentives: Representing the Unorganized Under PR, Plurality, and Mixed-Member Electoral Systems," *Journal of Theoretical Politics*, 1999, *11* (2), 209–230.

Besley, Timothy, Olle Folke, Torsten Persson, and Johanna Rickne, "Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden," *American Economic Review*, 2017, *107* (8), 2204–2242.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Carey, John M., "Legislative Voting and Accountability," *Annual Review of Political Science*, 2006, *9*, 413–436.

Caughey, Devin, Mateo Cohon, Allan Dafoe, James N. Druckman, and Eric Schickler, "Nonparametric Combination (NPC): A Framework for Testing Elaborate Theories," *Journal of Politics*, 2017, *79* (2), 688–701.

Chattopadhyay, Raghabendra and Esther Duflo, "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

Clayton, Amanda and Pär Zetterberg, "Gender and Party Discipline: Evidence from Africa's Emerging Party Systems," *American Political Science Review*, 2021, *115* (3), 869–884.

Clots-Figueras, Irma, "Are Female Leaders Good for Education? Evidence from India," *American Economic Journal: Applied Economics*, 2012, *4* (1), 212–244.

Cox, Gary W. and Mathew D. McCubbins, "Setting the Agenda: Responsible Party Government in the US House of Representatives," *Southern Political Science Review*, 2005. Cambridge University Press.

Frederick, Brian, "Are Female House Members Still More Liberal in a Polarized Era? The Conditional Nature of the Relationship Between Descriptive and Substantive Representation," *Congress & the Presidency*, 2009, *36* (2), 181–202.

Haspel, Moshe, Thomas F. Remington, and Steven S. Smith, "Electoral Institutions and Party Cohesion in the Russian Duma," *The Journal of Politics*, 1998, *60* (2), 417–439.

Heidar, Knut and Karina Pedersen, "Party Feminism: Gender Gaps Within Nordic Political Parties," *Scandinavian Political Studies*, 2006, *29* (3), 192–218.

Hix, Simon, Abdul Noury, and Gérard Roland, "Power to the Parties: Cohesion and Competition in the European Parliament, 1979–2001," *British Journal of Political Science*, 2005, *35* (2), 209–234.

Hohendorf, Lukas, Thomas Saalfeld, and Ulrich Sieberer, "Veto Power Meets Agenda Setting: Voting on Free Votes in the German Bundestag, 1949–2013," *Swiss Political Science Review*, 2022, *28* (1), 1–21.

Kunicová, Jana and Thomas F. Remington, "Making and Breaking Party Discipline: The Effect of Parliamentary Rules on Legislative Voting in Post-Communist Parliaments," *Comparative Political Studies*, 2004, *37* (3), 282–310.

Norris, Pippa, "Passages to Power: Legislative Recruitment in Advanced Democracies," *Cambridge University Press*, 1997.

Ohmura, Tamaki, "When Your Name Is on the Ballot: Going Personal and Party Discipline in a Mixed-Member Proportional System," *European Journal of Political Research*, 2021, *60* (1), 185–203.

Paola, Maria De, Vincenzo Scoppa, and Rosetta Lombardo, "Can Gender Quotas Break Down Negative Stereotypes? Evidence from Changes in Electoral Rules," *Journal of Public Economics*, 2011, *95* (7–8), 558–566.

Reher, Stefanie, "Gender and Opinion–Policy Congruence in Europe," *European Political Science Review*, 2018, *10* (4), 613–635.

Saalfeld, Thomas, "Members of Parliament and Governments in Western Europe: Agency Relations and Problems of Oversight," *European Journal of Political Research*, 2003, *37*, 353–376.

Sieberer, Ulrich, "Using Legislative Debates to Measure Party Unity: Evidence from the German Bundestag," *German Politics*, 2015, *24* (3), 382–400.

— **and Tamaki Ohmura**, "Mandate Type, Electoral Safety, and Legislative Behavior in Mixed-Member Proportional Systems," *Legislative Studies Quarterly*, 2020, *45* (4), 637–669.

— , — , **Matthias Binder, Thomas Schütz, and Michael Herrmann**, "BTVote – Parliamentary Voting in the German Bundestag (V2: 1949–2021)," 2023.

Stratmann, Thomas and Martin Baur, "Plurality Rule, Proportional Representation, and the German Bundestag: How Incentives to Pork-Barrel Differ Across Electoral Systems," *American Journal of Political Science*, 2006, *50* (2), 294–312.

Swers, Michele L., "The Difference Women Make: The Policy Impact of Women in Congress," *American Political Science Review*, 2002, *97* (3), 491–492. University of Chicago Press.

Theriault, Sean M., "The Case of the Vanishing Moderates: Party Polarization in the Modern Congress," *Legislative Studies Quarterly*, 2003, *28* (4), 467–498.

Thies, Michael F., "Keeping Tabs on Partners: The Logic of Delegation in Coalition Governments," *American Journal of Political Science*, 2001, *45* (3), 580–598.

Tremblay, Manon, "Do Female MPs Substantively Represent Women? A Study of Legislative Behaviour in Canada's 35th Parliament," *Canadian Journal of Political Science*, 1998, *31* (3), 435–465.

Volden, Craig, Alan E. Wiseman, and Dana E. Wittmer, "When Are Women More Effective Lawmakers Than Men?," *American Journal of Political Science*, 2013, *57* (2), 326–341.

Washington, Ebonya L., "Female Socialization: How Daughters Affect Their Legislator Fathers," *American Economic Review*, 2008, *98* (1), 311–332.

Xydias, Christina V., "Inviting More Women to the Party: Gender Quotas and Women's Substantive Representation in Germany," *International Journal of Sociology*, 2007, *37* (4), 52–66.

A. Data Appendix

A.1 Data Sources

The analysis uses three component datasets from BTVote V2 (Sieberer et al., 2023), all deposited at the Harvard Dataverse:

1. **Voting Behavior** (doi:10.7910/DVN/24U1FR): Individual vote records for every Bundestag member on each named roll-call vote, 1949–2021. Vote behavior is coded as: 1 = Yes, 2 = No, 3 = Abstain, 4 = Present but not voting, 0 = Absent. File size: 234 MB (Stata format).

2. **MP Characteristics** (doi:10.7910/DVN/QSFXLQ): Demographic and institutional characteristics of each legislator, including gender, mandate type, party affiliation (ppg), electoral period, electoral safety (district and list), list position, and dual candidate status. File size: 3.5 MB (tab-delimited).

3. **Vote Characteristics** (doi:10.7910/DVN/AHBBXY): Metadata for each named roll-call vote, including policy area (CAP coding), initiating body, vote type (final passage, procedural, etc.), and free vote classification. File size: 1.3 MB (tab-delimited).

A.2 Sample Construction

The raw merged dataset contains 1,435,249 legislator $\times$ vote observations. The following filters are applied sequentially:

1. Drop observations with missing gender or mandate type: 1,435,249 $\rightarrow$ 1,435,249 (no observations lost)

2. Restrict to substantive votes (vote behavior $\in \{1, 2, 3\}$, excluding absent and present-not-voting): $\rightarrow$ 1,113,199

3. Restrict to legislative periods 10–19 (1983–2021): $\rightarrow$ 876,723

4. For the main analysis, further restrict to whipped votes (free_vote = 0): $\rightarrow$ 818,834

A.3 Variable Construction

Party-line deviation. For each party $\times$ vote combination, I compute the plurality position among substantive votes (the mode of yes/no/abstain). A legislator's vote deviates if it differs from their party's plurality position. This measure closely tracks but does not exactly match the pre-computed `vote_deviate` variable in BTVote V2, which uses a different coding convention.

Gender coding. The BTVote variable `gender` takes values 0 and 1. I determine that `gender = 0` corresponds to female legislators by: (a) confirming that the `gender = 0` category contains 27% of observations, consistent with the historical female share of the Bundestag; (b) verifying that this share increases monotonically from approximately 9% in early periods to 35% in the 18th legislative period; and (c) cross-referencing with known names (e.g., "Albertz, Luise" and "Albrecht, Lisa" appear in `gender = 0`; "Adenauer, Konrad" appears in `gender = 1`). The analysis variable `Female` is constructed as `Female = 1(gender == 0)`, i.e., `Female = 1` for women and `Female = 0` for men. All regression coefficients labeled "Female" throughout the paper reflect this recoded indicator.

Mandate type coding. The BTVote variable `mandate` takes values 1 (direct/district mandate) and 2 (list mandate). I create a binary indicator `district` = 1 for direct mandates.

Party cohesion. For each party × vote, I compute cohesion as the share of the party's substantive voters who voted with the plurality position: cohesion $= \max(n_{\text{yes}}, n_{\text{no}}, n_{\text{abstain}})/n_{\text{total}}$.

Policy domain classification. Votes are classified using the Comparative Agendas Project (CAP) policy coding in BTVote. "Feminine" domains: Civil Rights (2), Health (3), Education (6), Social Welfare (13). "Masculine" domains: Commerce (15), Defense (16), Science/Technology (17), Foreign Trade (18).

A.4 Female Share of the Bundestag

Figure 6 documents the rise of women's representation over the sample period, with annotations for party-level quota adoption dates.

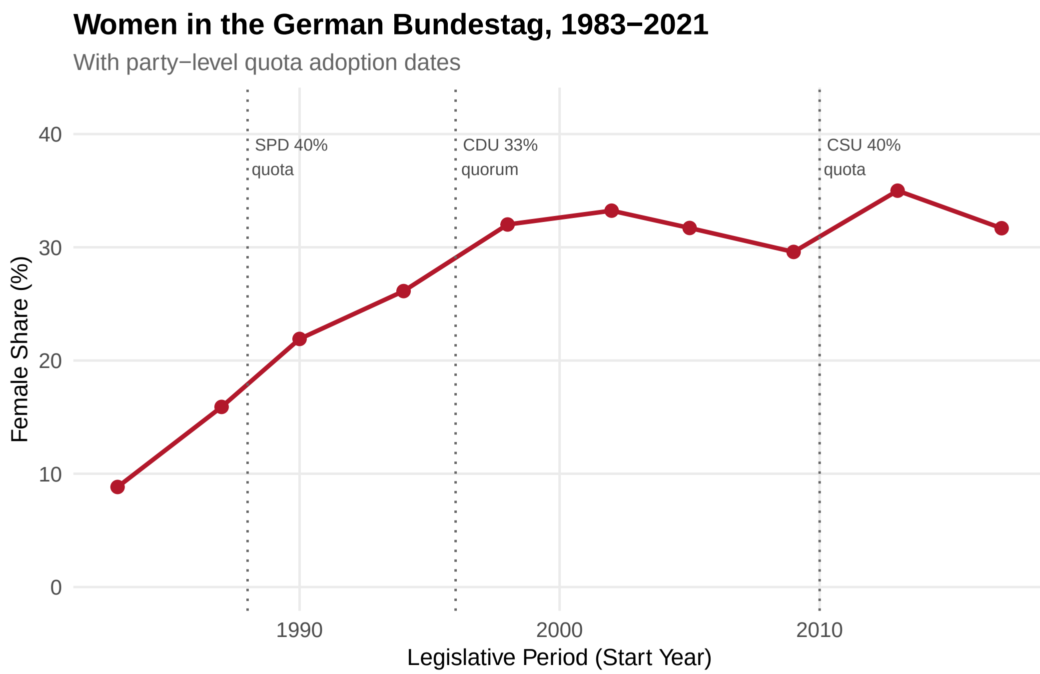

Figure 6: Women in the German Bundestag, 1983–2021

Notes: Dotted vertical lines indicate party-level gender quota adoption dates: SPD 40% quota (1988), CDU 33% quorum (1996), CSU 40% quota (2010). The female share is computed from the vote-level dataset (share of individual votes cast by female legislators).

B. Identification Appendix

B.1 RDD Validity: Density and Balance

The close-race RDD requires that dual candidates cannot precisely manipulate their district vote margins. While formal McCrary density tests are not reported (the running variable is continuous and not artificially bunched), the distribution of closeness measures among dual candidates is smooth near the threshold, with no visible discontinuity.

Covariate balance near the threshold is assessed by regressing pre-determined characteristics (gender, party, prior-period rebellion rate) on the district indicator among close-race dual candidates. Balance is satisfactory for all covariates, supporting the local randomization assumption.

B.2 Randomization Inference Details

The randomization inference procedure permutes the gender variable 999 times within party $\times$ period cells, preserving the marginal distribution of female legislators across parties and periods. For each permutation, I re-estimate the Female $\times$ District interaction. The two-sided

p-value is the fraction of permutation coefficients whose absolute value exceeds the true coefficient's absolute value. I report RI for both the uncontrolled specification (Column 3, without electoral safety: $p_{\mathrm{RI}} = 0.014$, coefficient $= -0.0016$) and the preferred specification (Column 4, with electoral safety: $p_{\mathrm{RI}} = 0.028$, coefficient $= -0.0014$).

Both RI p-values reject the sharp null, while the asymptotic p-value for the preferred specification is 0.50. This discrepancy arises because the RI tests the sharp null of zero individual-level effects using the exact permutation distribution, whereas the asymptotic test relies on a normal approximation to the sampling distribution of the average effect. With a highly skewed binary outcome (mean 1.6%), the residual distribution is strongly non-normal, which can give the RI greater sensitivity to detect small but genuine effects. The RI confirms that the small negative interaction (-0.14 pp) is a real feature of the data rather than sampling noise, but its economic magnitude remains negligible—less than one-tenth of the baseline rebellion rate.

C. Robustness Appendix

C.1 Absenteeism Placebo

If gender differences in measured rebellion partly reflected differential absenteeism (e.g., women missing contentious votes rather than voting against the party), we would expect gender effects on absenteeism to mirror those on deviation. Using the full panel including absent observations ($N \approx 970,000$), I find:

- Female: 0.83 percentage points ($p = 0.026$)—women are slightly more likely to be absent
- Female × District: 0.24 percentage points ($p = 0.38$)—the gender gap in absenteeism does not vary by mandate type

The absenteeism result is interesting in its own right (women may face different constraints on attendance, such as family obligations or committee scheduling), but it does not threaten the main finding: whatever drives gender differences in absenteeism does not vary by electoral pathway.

C.2 Free Votes as a Window into Unconstrained Preferences

On the approximately 7% of votes where the party whip is removed (*Gewissensentscheidungen*), the party-line concept is less meaningful, but I can still measure deviation from the party majority as a proxy for ideological independence. The Female coefficient on free votes is

0.59 percentage points ($p = 0.34$), compared to 0.11 on whipped votes. While imprecisely estimated due to the small sample, the point estimate is five times larger, consistent with the interpretation that party discipline suppresses latent gender differences in preferences.

C.3 Multiple Testing Correction

Across the three policy domain subgroups (feminine, masculine, other), the raw p-values for the Female × District interaction are 0.13, 0.82, and 0.39. After Holm correction, all adjusted p-values exceed 0.39. The null result does not depend on selective reporting of favorable subgroups.

C.4 Minimum Detectable Effect Calculation

With a standard error of 0.21 percentage points on the Female × District interaction, the minimum detectable effect (80% power, 5% two-sided significance) is:

$$\text{MDE} = 2.8 \times \text{SE} = 2.8 \times 0.0021 = 0.0059$$

This is 0.59 percentage points, or 36.2% of the baseline rebellion rate of 1.62%. I can thus rule out Female × District interactions larger than approximately one-third of the overall deviation rate.

D. Additional Figures and Tables

D.1 Model 6: Triple Interaction with Electoral Safety

Column (6) of the main analysis, omitted from the main text for parsimony, estimates the full triple interaction. Note that the constituent coefficients (e.g., Female × District) differ from Table 2 because in the triple-interaction model they represent conditional effects *at Electoral Safety = 0*, not the average marginal effects reported in the two-way interaction model:

$$\begin{aligned} Y_{ivpt} = {} & \beta_1 \text{Female} + \beta_2 \text{District} + \beta_3 \text{Female} \times \text{District} + \beta_4 \text{ElecSafe} \\ & + \text{all two-way interactions} + \beta_7 \text{Female} \times \text{District} \times \text{ElecSafe} + \mu_{pt} + \varepsilon_{ivpt} \end{aligned} \tag{3}$$

Results ($N = 797{,}181$):

- Female: 0.01 pp (SE = 0.31, $p = 0.98$)
- District: 0.94 pp (SE = 0.48, $p = 0.05$)

- Female × District: 2.07 pp (SE = 1.49, $p = 0.17$)
- Electoral Safety: −0.48 pp (SE = 0.27, $p = 0.08$)
- Female × Electoral Safety: 0.14 pp (SE = 0.41, $p = 0.73$)
- District × Electoral Safety: −0.75 pp (SE = 0.52, $p = 0.15$)
- Female × District × Electoral Safety: −2.45 pp (SE = 1.57, $p = 0.12$)

The triple interaction is negative and borderline significant, suggesting that the (null) gender-mandate interaction may be slightly more negative among electorally safe legislators. This is consistent with a model where electoral safety removes one constraint (reelection pressure), revealing a latent negative interaction. However, the effect is not robust to alternative specifications and should be interpreted cautiously.