Do Partisanship and Policy Agreement Make Citizens Tolerate Undemocratic Behavior?*

Kristian Vrede Skaaning Frederiksen**

October 19, 2021

Abstract

Do voters tolerate undemocratic behavior by political leaders with whom they share partisanship or policy interests? In this article, I answer this question employing candidate choice experiments in five democracies: the US, the UK, the Czech Republic, Mexico, and South Korea. In all five countries, I find that the negative effects of undemocratic behavior on voting intentions are at least as large for in-partisan candidates as for other candidates. Moreover, voters sanction candidates with whom they agree on policy two to three times as much as candidates with whom they disagree on policy. In line with prior studies, I also find that citizens tend to prefer undemocratic in-partisan candidates over democratically compliant out-partisan candidates but urge for methodological caution in that respect. I highlight that the more pessimistic conclusions of prior studies hinge on such a focus on *levels* of support for undemocratic and democratic candidates across partisanship and policy interests, which — in contrast to the focus on conditional *effects* of undemocratic behavior in this study — is rather sensitive to the strength of experimental treatments. Most importantly, this article shows that citizens are not blinded by their political interests when they face political leaders violating democratic principles.

*Word count: 7,518 (abstract: 199)

**PhD Candidate, Department of Political Science, Aarhus University

1 Introduction

Rejecting undemocratic political actors at the ballot box is seen as one of the most important tasks of ordinary citizens in order for democracies to survive (e.g., Svolik 2019: 29; Bermeo 2003: 8; Dahl 1971: 105; Lipset 1960: 87; Linz 1978: 18). The rise of authoritarian leadership in, for example, Hungary, Turkey, Venezuela, Poland, Brazil, and to some extent the United States through fairly democratic elections thus provide one of the greatest puzzles to contemporary political science (Levitsky and Ziblatt 2018: 88; Merkel 2014: 21; Stefes and Sehring 2011: 242). In fact, most democratic breakdowns today stem from subversions of democracy by political leaders elected by citizens (Svolik 2019). On the basis of extant accounts on how political priors color evaluations of actions and behaviors (e.g., Kunda 1990; Gaines et al. 2007; Lodge and Taber 2013; Tworzecki and Markowski 2014; Shapiro and Bloch-Elkon 2008), a likely explanation for popular support for undemocratic political leaders is that citizens' partisan interests make them turn the blind eye when politically favorable actors violate democratic principles. In this article, I provide answers to an important question in that regard: Do citizens simply tolerate undemocratic behavior by political leaders with whom they share partisanship or policy interests?

Some attention has already been devoted to the extent to which partisanship and policy interests drive citizens toward voting for undemocratic actors. A range of studies show that citizens are biased by their partisan priors when they evaluate violations of democratic principles (Carey et al. 2019; Ahlquist et al. 2018; Bartels 2020; Touchton et al. 2020; Albertus and Grossman 2021), but these studies do not examine whether these biases transmit onto voting behavior. Meanwhile, other studies have shown that citizens are willing to trade off democratic principles for policy interests and partisan loyalty in voting scenarios (Graham and Svolik 2020; Svolik 2020; Carey et al. 2020). However, these conclusions are drawn from the finding that voters, on the absolute level, show higher support for incrementally undemocratic in-partisan leaders as compared to democratically compliant out-partisan leaders. Thus, we still lack comprehensive evidence on the extent to which partisanship and policy agreement between voters and political

leaders suppress sanctioning, understood as negative effects, of undemocratic behavior when people vote.

To fill this gap, I provide experimental evidence from five democracies focusing on the extent to which partisanship and policy interests suppress the negative impact of undemocratic behavior on voting behavior. Specifically, I conducted well-powered candidate choice experiments in the United States, the United Kingdom, Mexico, South Korea, and the Czech Republic yielding a sample with more than 13,000 respondents and more than 230,000 observations. These candidate choice experiments enable me to examine the respective impacts of undemocratic behavior, policy agreement, and partisanship. Most importantly, the experiments make it possible to examine whether policy agreement and partisanship interfere with the impact of undemocratic behavior and thus to assess the extent to which citizens are biased by their partisan priors when they face undemocratic political leaders in voting scenarios. Moreover, the diverse case selection allows me to generalize the results broadly and advances the comparative scope of the literature substantially by providing one of the most comprehensive data collections on the link between undemocratic behavior and voting behavior to this date.

In all five countries, I find that the negative effects of undemocratic behavior on voting intentions are at least as large for in-partisan candidates as for other candidates. Moreover, voters sanction candidates with whom they agree on policy two to three times as much as candidates with whom they disagree on policy. Thus, I demonstrate that even though citizens tend to be biased by their partisan priors when they evaluate violations of democratic principles (Carey et al. 2019; Ahlquist et al. 2018; Bartels 2020; Touchton et al. 2020; Albertus and Grossman 2021), these biases do not travel to voting behavior. Transgressions against democracy is, therefore, an area where the impact of partisan priors on political behavior is rather limited. Meanwhile, I show that the findings from studies targeting voting (Graham and Svolik 2020; Svolik 2020; Carey et al. 2020) hinge on a focus on absolute levels of support for undemocratic and democratic candidates. In this five-country study, shifting the analytical focus from conditional effects to absolute levels also suggests that citizens prefer undemocratic in-partisans over democratically

compliant out-partisans. But I urge for caution when interpreting this finding: Focusing on levels of support for undemocratic and democratic candidates across political interests — as opposed to focusing on how political interests moderate the effects of undemocratic behavior — is rather sensitive to the relative strengths of experimental treatments representing partisanship, policy agreement, and undemocratic behaviors.

Most importantly, the findings show that citizens are not blinded by partisan loyalty and policy interests when they face undemocratic political leaders. This means that there is room for the support for undemocratic actors to erode over time if they continuously violate democratic principles or if these violations turn more severe.

2 Political Biases in Reactions to Undemocratic Behavior

I focus on undemocratic behavior as violations of the democratic cornerstones of free and fair elections, civil liberties, and/or the rule of law (Møller and Skaaning 2013). This definition captures the violations of democratic principles conducted by political leaders in the countries mentioned in the Introduction. These undemocratic actions have entailed manipulating the arrangement of courts and judges to the incumbent's benefit (e.g., Fidesz led by Prime Minister Victor Orbán in Hungary and the Law and Justice Party led by Prime Minister Jarosław Kaczyński in Poland), purging disloyal public officials (Presidents Recep Tayyip Erdoğan in Turkey and Donald Trump in the United States), harassing critics of the government (e.g., Presidents Jair Bolsonaro in Brazil and Andrés Manuel López Obrador in Mexico), and marginalizing the opposition's chances of reelection (e.g., Hugo Chávez and Nicolás Maduro in Venezuela) (e.g., Levitsky and Ziblatt 2018; Albertus and Grossman 2021).

The central question of this article is whether citizens respond differently to these behaviors depending on their partisanship and policy interests and, most importantly, whether in-partisans do not sanction political leaders for violating democratic principles when they vote. On the basis of studies investigating how citizens' political priors color

evaluation of information (e.g., Kunda 1990; Gaines et al. 2007; Lodge and Taber 2013; Tworzecki and Markowski 2014; Shapiro and Bloch-Elkon 2008), we may expect co-partisanship and alignment in policy positions between leaders and citizens to affect whether undemocratic behavior is perceived as actually undemocratic or, more generally, 'bad' or 'wrong'. A range of studies have already examined how partisanship and policy positions color citizens' interpretations of undemocratic behaviors, often concluding that these interpretations favor in-partisan and policy-proximate political leaders, candidates, or proposals (Krishnarajan, n.d.; Carey et al. 2019; Ahlquist et al. 2018; Bartels 2020; Touchton et al. 2020; Albertus and Grossman 2021; Beaulieu 2014).

What we do not know from these studies, however, is whether these biases transmit onto voting. Contrarily, Graham and Svolik (2020), Svolik (2020), and Carey et al. (2020) do examine the vote but put their theoretical emphasis differently than the stream of research mentioned above. These studies instead theorize on the extent to which citizens are willing to 'trade off' democratic principles for partisan loyalty and policy interests. The crux of partisan trade-offs when facing undemocratic political actors is that voters fail to "reveal a preference for democratic principles over other valid but potentially conflicting considerations such as political ideology, partisan loyalty, or policy preferences" (Graham and Svolik 2020: 392). This perspective — at least judging from Svolik (2020) and Graham and Svolik (2020) — paints a picture of a rational voter weighing off different considerations and letting the sum of these considerations decide who to vote for. Moreover, this perspective theorizes on overall levels of support — or vote shares — for undemocratic and democratically compliant leaders rather than on the conditional effects of undemocratic behavior across political interests. Voting for an undemocratic option that nevertheless aligns with the partisan loyalty or policy positions of the voter implies that the voter is willing to trade off democratic compliance for these other interests. Supporting the proposition that voters are willing to make this trade-off, Carey et al. (2020) as well as Graham and Svolik (2020) and Svolik (2020) find that citizens, on the absolute level, tend to vote for politically favorable candidates regardless of democratic compliance, also if these candidates violate several democratic norms (see

Carey et al. 2020: 9 for this latter finding).

Graham and Svolik (2020) and Carey et al. (2020) proceed to also testing whether co-partisanship interfere with the effects of undemocratic behavior in the United States, but the findings are mixed. Although conditional effects do not seem to be the main focus of the study, Graham and Svolik (2020: 400) with some statistical uncertainty find that strong partisans are less likely to sanction candidates from their own party as compared to candidates from the opposing party. Carey et al. (2020), meanwhile, find no differences in how voters respond to undemocratic behaviors by in-partisan and out-partisan candidates. These mixed findings coupled with an exclusive focus on the United States leave room for the focus of this article: That is, to concentrate on examining whether partisanship and policy interests diminish the negative effect of undemocratic behavior on the vote and to do this outside as well as in the United States.

The broader proposition I test in this article is, therefore, that partisanship and policy interests interfere with how voters sanction undemocratic behavior. Specifically, I expect this to be the case because voters interpret even factual information — including information about undemocratic behavior (Krishnarajan, n.d.; Carey et al. 2019; Ahlquist et al. 2018; Bartels 2020; Touchton et al. 2020; Albertus and Grossman 2021; Beaulieu 2014) — in ways consistent with their political priors (Kunda 1990; Gaines et al. 2007; Lodge and Taber 2013; Tworzecki and Markowski 2014; Shapiro and Bloch-Elkon 2008).

I examine three pre-registered hypotheses derived from this proposition. The first is that partisanship — negative as well as positive — diminishes the negative effect of undemocratic behavior. I expect negative partisanship (i.e., out-partisanship) as well as positive partisanship (i.e., in-partisanship) to limit sanctions for undemocratic behavior because support for out-partisan political leaders plausibly is floored regardless of undemocratic behavior. Thus, it is when citizens hold neutral feelings toward the particular leader's party that I expect the strongest negative effect of undemocratic behavior.

The two remaining hypotheses concern policy agreement between citizens and political leaders. I distinguish here between policy agreement in terms of *distance* and *intensity*, the former being the promixity between positions and the latter being the in-

Table 1: Hypotheses

Partisanship:
Partisanship — negative as well as positive — diminishes the negative effect of undemocratic behavior.
Policy Interests:
Large policy distance or absence of policy distance to the candidate diminish the negative effect of undemocratic behavior
Intense policy preferences — to the same or opposite side of the candidate — diminish the negative effect of undemocratic behavior

tensity of a position either aligned or opposed to the leader's (Sartori 1976: 111). I simply do this to examine whether the two types of policy agreement matter differently for how citizens sanction undemocratic behavior. Thus, I expect large policy distance or absence of policy distance to the candidate to diminish the negative effect of undemocratic behavior, and I expect intense policy preferences — to the same or opposite side of the candidate — to diminish the negative effect of undemocratic behavior. As on partisanship, I expect U-shaped relationships because support for candidates with whom citizens disagree profoundly on policy is likely to be so low that there is no room for an effect of undemocratic behavior. It is when the voter neither profoundly disagrees nor agrees with the candidate on policy that I expect the largest negative effect of undemocratic behavior. Table 1 sums up the three hypotheses.

3 Research Design and Data

I employ well-powered candidate choice experiments fielded via Lucid in September and October 2020 in the United States, the United Kingdom, Mexico, South Korea, and the Czech Republic to examine whether partisanship and policy interests interfere with how voters respond to undemocratic behavior (N = 2,350-2,999 respondents and n = 41,788-53,417 candidates in each country).[1] The experiments make this examination possible by manipulating candidate partisanship, policy positions, and undemocratic behaviors. The

[1]See Coppock and McClellan (2019) for a general examination of the performance of Lucid-samples showing that these typically resemble national election studies fairly well. The samples employ here at nationally representative on age and gender.

experiments and analysis plans are pre-registered.[2] Deviations from the pre-registration are described in Appendix L.[3]

The selected countries vary on parameters suited to generalize the findings of this article. First, there is substantial variation in polarization and the intensity of party identification. Specifically, the United States represents a case of high polarization and intense party identification while ideological rigidity and partisan identifications are particularly weak in Mexico (Langston 2017: 112), modest in South Korea and in the Czech Republic (Lee 2016: 163; Hajek 2017: 285), and somewhat increasing in recent years in Britain (Hobolt et al. 2020). Second, there is substantial variation in democratic and autocratic legacies between the countries which is important to maximize variation in prior experiences with democratic compliance. Specifically, the United Kingdom and the United States are very old democracies while South Korea (prior military dictatorship), the Czech Republic (prior Communist dictatorship under Soviet influence), and particularly Mexico (prior hegemonic one-party rule) are younger democracies (Kang 2017: 16; Langston 2017: 1). Finally, two of the countries have suffered recent attacks on democracy — the United States and Mexico — while the three remaining countries, except for corruption scandals, have remained clear of democratic backsliding. This all mean that common tendencies in the prevalence of political biases in how voters respond to undemocratic behavior would be strongly generalizable.

3.1 Candidate Profiles

The experiments presented each respondent for 10 pairs of candidates (totalling 20 candidates) contesting for the presidency in the United States, the Czech Republic[4], Mexico, and South Korea, and for prime minister in the United Kingdom. The respondents were asked to state their likelihood to vote for (to support in the UK) each of the candidates on a five-point scale from "Very unlikely" to "Very likely". The candidates vary

[2] I test H3a-c from the pre-registration in this letter. The remainder of the hypotheses will — due to conciseness concerns — be reported elsewhere.

[3] The pre-registration is public and can be found here (for this presentation only — not for anonymous review): https://osf.io/qjm42.

[4] Although the powers of Czech presidents are more limited than in the remaining countries, Czech presidents do still have considerable political powers.

on background attributes, skills in fighting corruption and in handling economic matters[5], partisanship, policy positions, and undemocratic behavior. All attributes, their assignment probabilities, ethical considerations, and a candidate scenario are included in Appendix A.

I assigned background attributes (i.e., age, gender, and background) and their probabilities on the basis of substantive knowledge of real-world distributions of current Mexican and American governors, prior South Korean and Czech presidential candidates, and former British prime ministers to enhance the external validity of the experiments (de la Cuesta et al. 2021). The candidates that the respondents face are, therefore, similar to candidates that they would encounter in the real world. Attribute levels on the dimensions of theoretical interest (i.e., partisanship, policy positions, and undemocratic/democratic behavior), on the other hand, are randomized with equal probabilities.[6] I assigned policy positions across the issue dimensions of redistribution and morality policy. The positions on redistribution consist of realistic proposals on tax policy, welfare spending, education, and labor unions, while the positions on morality policy include proposals on abortion, immigration, crime, and same-sex marriage (see also Table A1-A3 in Appendix A).

I assigned undemocratic behavior as violations of the democratic cornerstones of free and fair elections, civil liberties, and the rule of law as shown in Table 1.[7] Each candidate was assigned either a democratically compliant or an undemocratic behavior. These violations are incremental and piecemeal to realistically reflect how they play out in contemporary democracies (e.g., Graham and Svolik 2020; Levitsky and Ziblatt 2018). Moreover, consistent with prior studies the wordings of undemocratic behaviors are neutral rather than presented in a leading, negative language (Graham and Svolik 2020: 397; Carey et al. 2020: 4). This means that I isolate the effects of undemocratic behavior

[5]This dimension was manipulated with the purpose of an empirical test not included here and is, therefore, not mentioned in the remainder of the article. The attribute levels within this dimension were assigned randomly and are, therefore, unlikely to bias the results.

[6]This obviously deviates from the real-world distribution of these attributes: For example, policy positions and partisanship are not distributed independently of each other in the real world. However, randomizing these factors independently of each other is the most fruitful choice because I aim to examine the independent effects of each of these factors.

[7]What I term 'rule of law' may be defined as 'checks and balances' elsewhere (e.g., Graham and Svolik 2020, 397).

Table 2: Undemocratic (top) and Democratic (bottom) Behaviors

Electoral fairness: Supported a proposal to reduce polling stations in areas that support opposing parties
Rule of Law: Said court rulings by judges appointed by opposing parties should be ignored
Civil Liberties: Said it is legitimate to fight political opponents in the streets if one feels provoked Said it is acceptable to harass journalists that do not reveal sources
Electoral fairness: Supported a proposal to preserve existing polling-stations in all areas
Rule of Law: Said court rulings by judges appointed by opposing parties should be adhered to
Civil Liberties: Said it is unacceptable to fight political opponents in the streets even though one feels provoked Said it is unacceptable to harass journalists even though they do not reveal sources

rather than capture general positive or negative valence of the statements. This setup, which is similar to prior studies focused on levels of support for undemocratic and democratic candidates (Graham and Svolik 2020; Svolik 2020), provides an easy test for the proposition that voters are willing to trade off democratic compliance for partisan ends because the violations of undemocratic behavior are in fact incremental, piecemeal, and neutral (i.e., they are not "strong" treatments). These rather soft violations therefore also provide a fairly easy test for the proposition that political interests interfere with the negative effects of undemocratic behavior, which is central to this article, as they provide fertile ground for partisan interpretations.

3.2 Measuring Shared Partisanship and Policy Agreement

I measured shared partisanship and policy agreement — as policy distance and as directional intensity — between respondents and candidates to test the three hypotheses of this article. Before they encountered any candidates, I measured partisanship as the

respondents' attitudes to each of the major political parties in their country[8] on a five-point scale ranging from "Dislike a great deal" to "Like a great deal". Afterwards I matched this information with the party of each of the candidates attaining a measure of the respondents' affect toward the specific party of the assigned candidate.[9] Affective like/dislike-scores are useful here as they yield a comparatively consistent measure of shared candidate-respondent partisanship across contexts and capture negative as well as positive partisanship (Abramowitz and Webster 2018; Wagner 2021).

I conducted the measure of policy distance by combining information about the respondents' political preferences with the policy positions assigned to the candidates. First, I measured policy preferences by the respondents' attitudes toward a series of statements on a five-point scale ranging from "Strongly disagree" to "Strongly agree". These statements, which are shown in Table A4 in the appendix, touch on the exact same issues as the positions assigned to the candidates. Second, I flipped the scales so that they all range from an extreme left-wing position to an extreme right-wing position and matched this information with the specific issue positions of the candidates.[10] Finally, I computed the total policy distance between respondents and candidates across redistribution and morality issues attaining a measure ranging from 1 to 7 where low values signal little to no policy distance and high values signal large policy distance.[11][12]

I conducted the measure of preference intensity to the same or opposite side of the candidate by combining information from the agreement statements mentioned above with another battery asking the respondents to place each of the issues on a five-point

[8]These parties include PRI, PAN, MORENA, and PRD in Mexico, UFP and DPK in South Korea, ANO 2011, ODS, and ČSSD in the Czech Republic, Labour and Conservatives in the UK, and the Democrats and Republicans in the US.

[9]I also asked the respondents whether they identify with a specific party (and if so, which party). However, I only use this measure as a supplement.

[10]For example, if the candidate within the morality policy dimension was assigned a position on abortion, I let the respondent's preference on abortion represent preference extremity on morality issues on that specific observation. If the respondent then encounters another candidate that takes on a position on same-sex marriage, I let the respondent's preference on same-sex marriage represent preference extremity on morality issues on that next observation.

[11]Left-wing candidate positions were assigned a value of 1.5 on the five-point scale whereas right-wing positions were assigned a value of 4.5. I also created supplementary measures using values slightly above and below this to check the sensitivity to this decision (see also Appendix E).

[12]Specifically, the measure captures the summed numerical distance across the issues. For example, a distance of 7 signals that the distance from candidate to respondent is 3.5 on both issues (either by 1.5 against 5 or 4.5 against 1).

scale from "Not at all important" to "Extremely important". On the basis of the agreement statements, I first coded whether the respondents took on positions on the same or opposite wing of the candidate on the particular issues on which the candidate was assigned positions. Same-side positions were coded as 1 whereas opposing positions were coded as -1.[13] I then multiplied these values with the respondents' importance statements and summed across the dimensions of morality issues and redistribution, yielding a continuous measure ranging from -5 to 5 where -5 signals intense preferences to the opposite side of the candidate, -1 signals mild preferences to the opposite side of the candidate, 1 signals mild preferences in the same direction as the candidate, and 5 signals intense preferences in the same direction as the candidate.[14]

3.3 Model Specifications

I estimate the Average Marginal Compontent Effects (AMCEs) of undemocratic behavior conditioned by partisanship and the two measures of policy agreement — that is, interaction models — using linear regression with the candidate assessments as the dependent variable. I cluster standard errors on the respondent level to account for the fact that each respondent rates multiple candidates (Hainmueller et al. 2014). I treat partisanship as a factor variable (i.e., a categorical variable) as the interaction effects of this variable did not turn out to have any clear functional form. Consistent with the pre-registration, I include second-order terms on the policy agreement measures because polynomial interactions best fitted the data there.

The hypotheses receive support if the interaction effects turn out to be curve-linear: That is, if the negative effects of undemocratic behavior are smallest when the candidate is either very proximate or very distanced to the respondent on policy, when

[13] I assigned a value of -0.5 if the respondent took on a completely neutral position on the particular issue that the candidate took either a left- or right-wing position on. I exclude these respondents as a robustness check in Appendix E, showing that this choice does not make any difference to the results.

[14] Note that values close to zero (e.g., -1 or 1) also can signal intense preferences that conflict across the two policy dimensions. That is, low values can also signal that the respondent holds an intense preference against the candidate on morality issues and holds an intense preference in the same direction as the candidate on redistribution or vice versa. Disaggregating the measure into its two individual dimensions removes this feature, and I show in Appendix C that such disaggregation does not change the results shown in the empirical analysis.

the respondent has intense preference to the same or opposite side of the candidate, and when the respondent holds either very negative or positive feelings toward the candidate's party. Conversely, I expect the effects to be largest when the respondent holds a neutral attitude toward the candidate's party or neither profoundly disagrees nor agrees with the candidate on policy (i.e., when the partisanship and policy measures are approximately at their mid-points).

4 Analysis

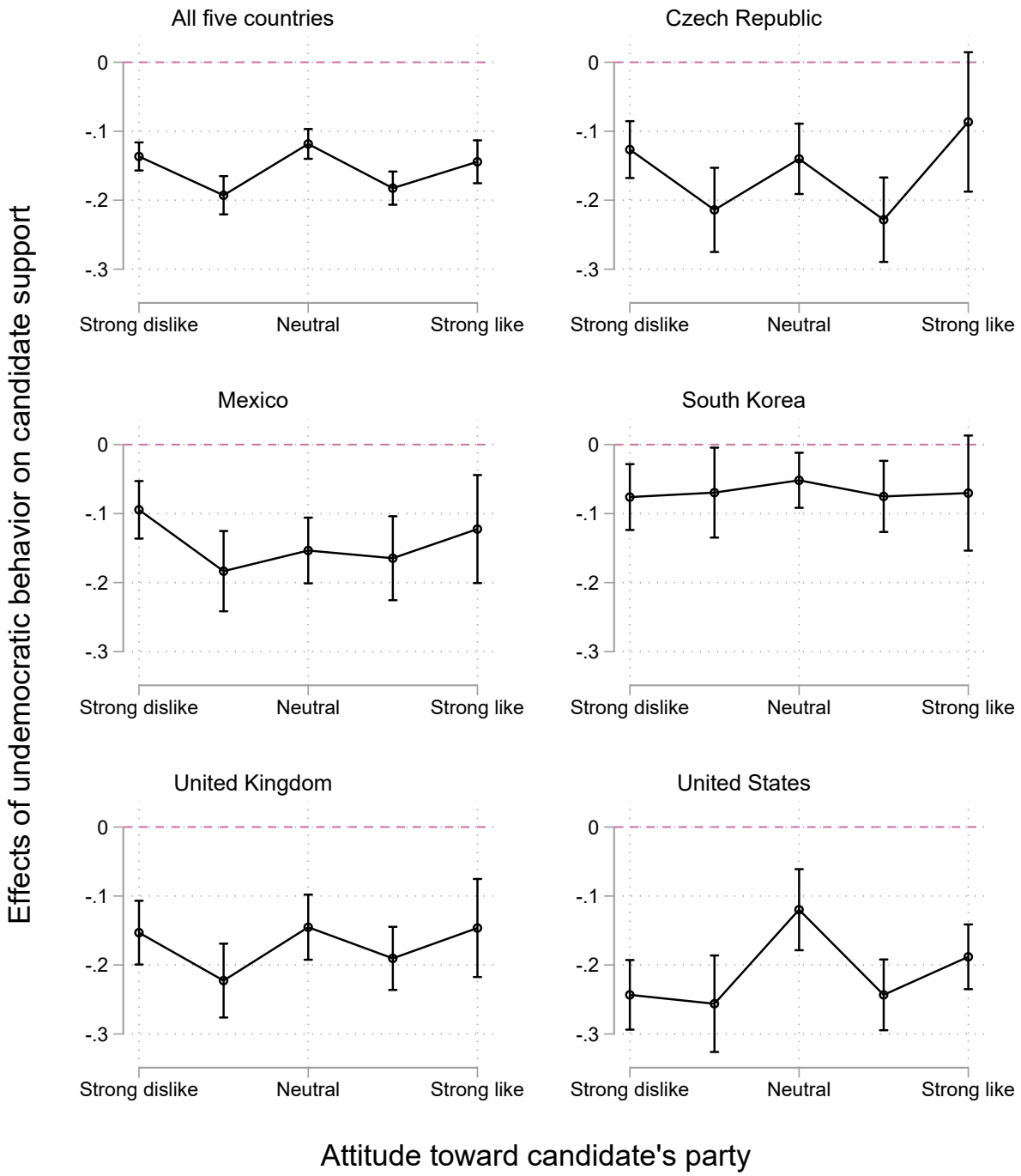
4.1 Do Partisan Loyalty and Policy Interests Interfere with Sanctions of Undemocratic Behavior?

Figures 1-3 illustrate the AMCEs of undemocratic behavior conditioned by partisanship and the two measures of policy agreement, allowing us to assess the extent to which partisan loyalty and policy interests interfere with how citizens sanction undemocratic behavior. Specifically, Figure 1 shows the effects of undemocratic behavior for those with strong negative or positive feelings toward the candidate's party at the endpoints, for those with neutral feelings toward the candidate's party in the middle, and for those with moderate positive or negative feelings toward the candidate's party between the endpoints and the middle. Similarly, Figures 2 shows the effects of undemocratic behavior across the policy distance measure increasing from 1 to 7. Finally, Figure 3 shows the effects of undemocratic behavior across the preference intensity measure with intense preferences for or against the candidate's positions at the endpoints and corresponding non-intense preferences[15] near the middle.

Starting with Figure 1, we should look for whether voters sanction undemocratic behavior less when they hold strong feelings toward the candidate's party. We see that this expectation does not hold: Figure 1 shows that the effects of undemocratic behavior within and across countries fluctuate unsystematically between just below -0.1 and just above -0.25 at the different levels of positive and negative partisanship. There is a small

[15] Or have cross-cutting intense preferences, cf. footnote 14.

Figure 1: Effects of undemocratic behavior on candidate support across shared partisanship between voter and candidate.



Note: OLS-regression including an interaction term between undemocratic behavior and shared partisanship between respondent and candidate and with standard errors clustered on the respondent-level. 2,350-2,999 respondents and 41,788-53,417 candidates in each country (13,308 and 239,356 in the pooled estimate).

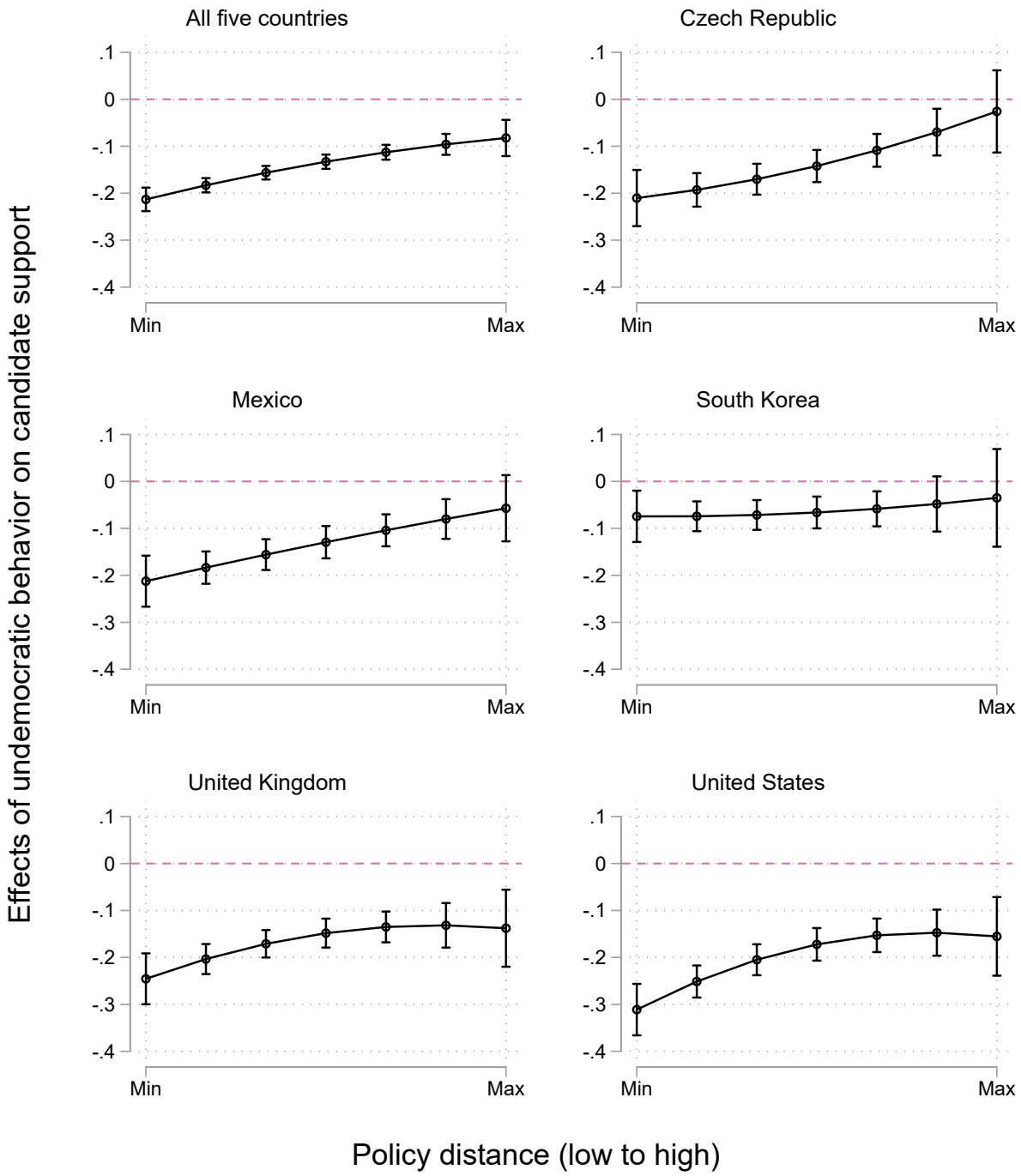
tendency — except in South Korea where partisanship seems to have no moderating impact at all — that undemocratic behavior is sanctioned the most if the voter has moderate either positive or negative feelings toward the candidate's party. In the pooled estimate, the effect of undemocratic behavior is larger for moderate partisans than for neutrals ($p < 0.001$) and somewhat larger for moderates than for strong partisans ($p < 0.1$), while neutrals and strong partisans show more similar and statistically indistinguishable effects. Interestingly so, in the United States it is when voters feel neutral toward a candidate's party that they sanction undemocratic behavior the least by around half as much as they sanction in-partisan and out-partisan candidates. This contrasts the finding that voters sanction in-partisan candidates less for behaving undemocratically found in Graham and Svolik (2020).

In Figures 2-3, we should look for to what extent voters sanction undemocratic behavior less when they either profoundly agree or disagree with the candidate on policy. Figure 2 provides evidence against this proposition. Instead, we see that the effect of undemocratic behavior decreases with the policy distance to the candidate. Specifically, voters in all countries except South Korea punish candidates with whom they profoundly agree (minimum distance) on policy two to three times as much as they punish candidates with whom they profoundly disagree (maximum distance). As was the case with partisanship, policy distance does not moderate the impact of undemocratic behavior in South Korea.

Figure 3 shows that the picture is much the same on directional intensity. In the United Kingdom, the United States, and Mexico, the impact of undemocratic behavior increases with the degree to which the voter has intense preferences in favor of the candidate. In the Czech Republic, undemocratic behavior is also sanctioned the most if voter and candidate agree on policy but the intensity of preferences only matters when they disagree. As was the case with partisanship and policy distance, preference intensity does not interfere with the effect of undemocratic behavior in South Korea.

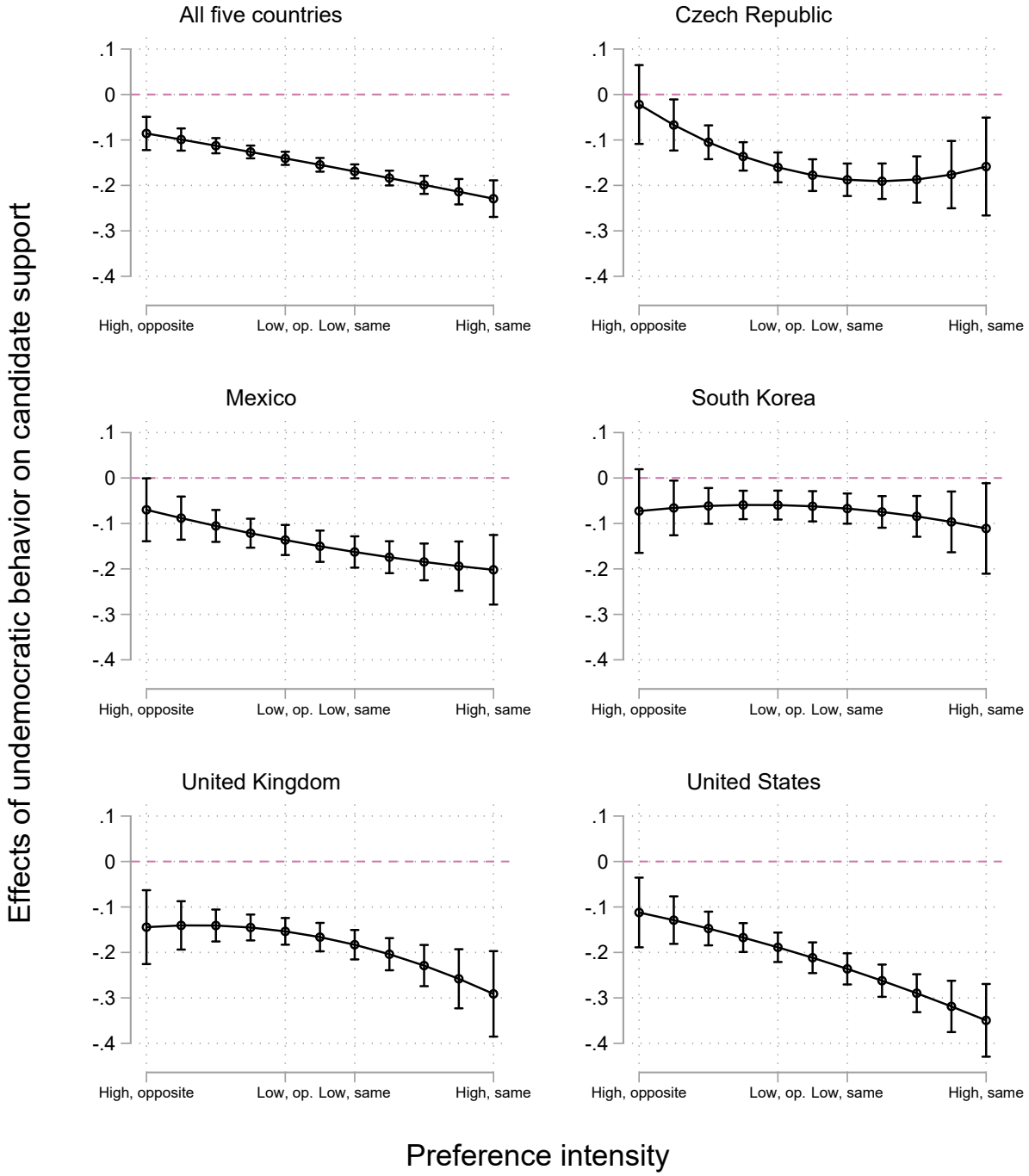
In sum, we reject all three hypotheses: Partisanship does not diminish the negative effect of undemocratic behavior and voters even sanction undemocratic candidates

Figure 2: Effects of undemocratic behavior on candidate support across policy distance between voter and candidate.



Note: OLS-regression including an interaction term between undemocratic behavior and squared policy distance between respondent and candidate and with standard errors clustered on the respondent-level. Same sample as Figure 1.

Figure 3: Effects of undemocratic behavior on candidate support across preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate.



Note: OLS-regression including an interaction term between undemocratic behavior and squared intensity in policy agreement between respondent and candidate and with standard errors clustered on the respondent-level. Same sample as Figure 1.

with whom they agree on policy the most.

4.2 Levels of Support for Undemocratic and Democratic Political Leaders

Prior studies have inferred that citizens are willing to trade off democratic compliance for partisan loyalty and policy interests from the finding that they prefer politically favorable undemocratic candidates over politically unfavorable democratic candidates (Svolik 2020; Graham and Svolik 2020). As mentioned earlier, these inferences are based on levels of support for undemocratic and democratic candidates rather than conditional effects of undemocratic behavior across political interests. But does this willingness to trade off democratic compliance for other political interests replicate in this five-country experiment?

Table 3: Levels of support for undemocratic and democratic candidates. 'Unfavorable' denotes either 'Strong dislike' on partisanship, maximum distance on policy distance, and maximum intense disagreement on preference intensity.

	Partisanship	Distance	Intensity
Politically Unfavorable Democratic	2.08	2.31	2.41
	(0.01)	(0.02)	(0.02)
Politically Favorable Undemocratic	3.37	2.93	3.11
	(0.02)	(0.01)	(0.02)

Standard errors in parentheses

Same sample and model specifications as Figures 1-3.

It indeed does. In Table 3, I have included the levels of support for the most politically unfavorable democratic candidates — that is, candidates with whom the voter disagrees the most on policy and candidates from parties that the voter strongly dislikes — versus the most politically favorable undemocratic candidates. Strongly in-partisan, undemocratic candidates defeat strongly out-partisan, democratic candidates by huge margins of approximately 1.3 scale points, equal to 32 percentage points, in the pooled estimate. Voters also prefer undemocratic candidates from parties they strongly like to democratic candidates from parties they feel neutral towards or like somewhat ($p < 0.001$

in both comparisons). On policy distance, voters prefer the most proximate but undemocratic candidates over the most distanced but democratically compliant candidates by approximately 0.6 scale point ($p < 0.001$). On directional preference intensity, voters prefer undemocratic candidates with whom they share intense preferences to democratically compliant candidates on the opposite wing with whom they disagree intensely on policy by approximately 0.7 scale point ($p < 0.001$; see also Appendix B for an elaborated graphic illustration of the levels of support for undemocratic and democratic candidates).

Judging from this experimental setting, voters therefore prefer politically favorable undemocratic candidates over politically unfavorable democratic candidates. We could, in line with prior studies, infer from this finding that citizens are willing to trade off democratic compliance for partisan ends. But I urge for caution before drawing this conclusion as it is driven by the individual effects of partisanship, policy agreement, and undemocratic behavior and, therefore, the particular experimental treatments used in this study. In this particular experimental setup, the effect of undemocratic behavior, which as shown in Figures 1 typically varies from -0.1 to -0.25 scale point, is tiny compared to partisanship and policy agreement across countries (see also Appendix I).

In sum, citizens prefer in-partisan candidates who violate democratic principles incrementally to out-partisan candidates who are democratically compliant. This is intuitively worrying from a democratic perspective. But we need to be cautious on how to interpret this finding as it depends heavily on the strength of experimental treatments. In this and other studies (Graham and Svolik 2020; Carey et al. 2020; Svolik 2020), the undemocratic behaviors assigned to candidates are incremental and neutral, and therefore provides an easy test of the proposition that voters are willing to trade off democratic compliance for partisan ends. Assigning more severe undemocratic behaviors or providing diagnostic information about them (e.g., stating that they are, in fact, undemocratic) might yield different results as the balance between experimental treatments would be altered.

4.3 Robustness Checks and Auxiliary Analyses

In the appendix, I show that splitting the undemocratic behavior variable in its four antagonistic pairs (i.e., supporting electoral manipulation, discarding judges appointed by opposing parties, encouraging violence, and endorsing journalist harassment) or splitting the policy agreement measure in its two dimensions (i.e., redistribution and morality issues) does not change the findings of Figures 1-3.

I also utilize the differing effect sizes of the individual undemocratic behavior-items to illustrate the sensitivity of the finding — outlined in Section 4.2 — that voters prefer politically favorable undemocratic candidates over politically unfavorable democratic candidates (i.e., sensitivity to the strength of experimental treatments). Moreover, I provide new data on another type of respondents — that is, American MTurkers — to assess both the sensitivity of the findings from Figures 1-3 and the findings outlined in Section 4.2. The findings from Figures 1-3 are robust to using this data while the sensitivity of the findings outlined in Section 4.2 are further exposed by this test, plausibly because MTurkers are more politically aware and thus perceive the particular undemocratic behaviors to be more severe than Lucid-respondents do (Coppock and McClellan 2019: 6).

Moreover, I show that the results from Figures 1-3 are robust to using an alternative, categorical — rather than affective — measure of partisanship and to adjusting the policy agreement measures slightly in terms of how extreme the candidates are assumed to be. I also provide tests with different parametric assumptions and show that the findings remain the same even if partisanship is used in its squared form or if the policy agreement measures are treated as factor variables. Furthermore, I show that including respondent covariates (gender, education, and urban/rural residence) or including the remaining candidate attributions in the specifications does not change the results. Finally, I provide supplementary tables as an alternative presentation of the results in Figures 1-3, supplementary figures to the results mentioned in Section 4.2, a standard AMCE-plot showing the average effects of the different candidate attributes, and a test of whether having extreme policy preferences in itself matters for sanctioning of undemocratic behavior.

5 Discussion and Conclusion

Are voters so biased by their political priors that they do not punish violations of democratic principles by their preferred parties and candidates? This question is central to understanding the roles of partisan loyalties and policy interests in democratic backsliding. Employing survey-experimental evidence from five contemporary but very different democracies, I provide an answer that is encouraging from a democratic perspective: Voters punish in-partisan candidates as much as other candidates for behaving undemocratically, and they punish candidates with whom they agree on policy more than candidates with whom they disagree.

The novelties of these findings are that they shed light on which effects undemocratic behavior has on voting for politically unfavorable and favorable leaders across different countries. First, the findings suggest that even though citizens may interpret undemocratic behavior and democratic principles in a biased manner (Carey et al. 2019; Ahlquist et al. 2018; Bartels 2020; Touchton et al. 2020; Albertus and Grossman 2021), these biases do not affect the vote. When it comes to responses to violations of democratic principles, the impact of political priors is therefore limited. Second, the findings show that the more discouraging conclusions on the democracy-distorting roles of partisan loyalty and policy interests of prior studies (Svolik 2019; Svolik 2020; Graham and Svolik 2020) hinge on a focus on absolute levels of support for undemocratic and democratic leaders across political interests. The latter point is especially important as such a focus on levels of support, rather than conditional effects, is sensitive to the relative strength of experimental treatments representing undemocratic behavior and political interests.

The findings of this article, however, also pose some questions looking forward. First, although the main finding that citizens are not biased by partisan priors when they sanction undemocratic behavior replicated across all five countries, we did still see some cross-country variation. For example, the impact of undemocratic behavior was particularly small in South Korea. South Korea was, moreover, the only country in which the most policy-proximate candidates were not punished the most for behaving undemocratically. Explaining such comparative differences in how citizens sanction undemocratic

behavior is beyond the scope of this article but may be a fruitful venue for future research. Another unanticipated finding was exactly that candidates with whom citizens agree on policy get sanctioned the most for undemocratic behavior. This finding resemble those of Reuter and Szakonyi (2019) who describe a similar result on electoral fraud in Russia as a "shock effect" where those who initially support President Putin have the highest *ex ante* expectations about the cleanness of elections. Analogously, one could argue here that voters have high expectations about democratic compliance among leaders that they agree with politically and therefore get "shocked" if these leaders behave undemocratically.

Finally, it was rather unexpected that policy interests and partisanship play different roles in how citizens sanction undemocratic behavior: Alignment between citizens and candidates increases sanctioning of undemocratic behavior on policy, but not on partisanship. This could suggest that voters have a harder time setting their partisan loyalties aside as compared to their policy interests. It could also suggest that "shock"-effects and partisan biases, which are contradicting mechanisms, are playing out when citizens face undemocratic in-partisan political leaders. Future research would benefit from examining why there is such a difference in how policy interests and partisanship moderate the negative effect of undemocratic behavior.

What broader implications do these findings, then, have for democratic sustainability? Because citizens are not biased by their political priors in how they sanction undemocratic behavior, there is indeed potential for vote-switching away from the most proximate (potentially undemocratic) political leaders and parties. Such vote-switching could play out if political leaders continue to violate democratic principles over time, because these violations turn more severe, or if citizens start to perceive even incremental undemocratic behavior as severe. Meanwhile, the rather small impacts of these incremental violations tell us that citizens need reasonable alternatives to be willing to switch their vote. The encouraging point of this article is that citizens are not blinded by their political interests when they face political leaders violating democratic principles.

References

Abramowitz, A. I. and S. W. Webster (2018). Negative partisanship: Why americans dislike parties but behave like rabid partisans. *Political Psychology 39*(S1), 119–135.

Ahlquist, J. S., N. Ichino, J. Wittenberg, and D. Ziblatt (2018). How do voters perceive changes to the rules of the game? Evidence from the 2014 Hungarian elections. *Journal of Comparative Economics 46*(4), 906 – 919.

Albertus, M. and G. Grossman (2021). The americas: When do voters support power grabs? *Journal of Democracy 32*(2), 116–31.

Bartels, L. M. (2020). Ethnic antagonism erodes republicans' commitment to democracy. *Proceedings of the National Academy of Sciences 117*(37), 22752–22759.

Beaulieu, E. (2014). From voter id to party id: How political parties affect perceptions of election fraud in the u.s. *Electoral Studies 35*, 24 – 32.

Bermeo, N. (2003). *Ordinary People in Extraordinary Times: The Citizenry and the Breakdown of Democracy.* Princeton, New Jersey: Princeton University Press.

Carey, J., K. Clayton, G. Helmke, B. Nyhan, M. Sanders, and S. Stokes (2020). Who will defend democracy? evaluating tradeoffs in candidate support among partisan donors and voters. *Journal of Elections, Public Opinion and Parties 0*(0), 1–16.

Carey, J. M., G. Helmke, B. Nyhan, M. Sanders, and S. Stokes (2019). Searching for bright lines in the trump presidency. *Perspectives on Politics 17*(3), 699–718.

Coppock, A. and O. A. McClellan (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research and Politics 6*(1), 1–14.

Dahl, R. A. (1971). *Polyarchy: Participation and Opposition.* New Haven: Yale University Press.

de la Cuesta, B., N. Egami, and K. Imai (2021). Improving the external validity of conjoint analysis: The essential role of profile distribution. *Political Analysis*, 1–27.

Gaines, B. J., J. H. Kuklinski, P. J. Quirk, B. Peyton, and J. Verkuilen (2007). Same facts, different interpretations: Partisan motivation and opinion on iraq. *Journal of Politics 69*(4), 957–974.

Graham, M. H. and M. W. Svolik (2020). Democracy in America? partisanship, polarization, and the robustness of support for democracy in the United States. *American Political Science Review 114*(2), 392–409.

Hainmueller, J., D. J. Hopkins, and T. Yamamoto (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis 22*(1), 1–30.

Hajek, L. (2017). Left, right, left, right... centre: Ideological position of andrej babiš's ano. *Politologický časopis - Czech Journal of Political Science 3*, 275–301.

Hobolt, S. B., T. J. Leeper, and J. Tilley (2020). Divided by the vote: Affective polarization in the wake of the brexit referendum. *British Journal of Political Science*, 1–18.

Kang, W. J. (2017). Income and voting behavior in korean politics: Why do the poor support conservative political parties? *Journal of International and Area Studies 24*(2), 15–33.

Krishnarajan, S. (n.d.). Rationalizing democracy: The partisan perceptual bias of undemocratic proposals. *Working paper*.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin 108*(3), 480–498.

Langston, J. K. (2017). The dinosaur that did not die: Mexico's pri from 1982 to 2012. In J. K. Langston (Ed.), *Democratization and Authoritarian Party Survival: Mexico's PRI*. New York: Oxford University Press.

Lee, J. (2016). *A Study on the Party System in South Korea after Democratization.* Ph. D. thesis, The University of Michigan.

Levitsky, S. and D. Ziblatt (2018). *How Democracies Die: What History Reveals About Our Future.* Penguin Books.

Linz, J. J. (1978). *The Breakdown of Democratic Regimes: Crisis, Breakdown and Reequilibration. An Introduction.* Johns Hopkins University Press.

Lipset, S. M. (1983[1960]). *Expanded Edition, Political Man: The Social Bases of Politics.* London: Heinemann.

Lodge, M. and C. S. Taber (2013). *The Rationalizing Voter.* New York, NY: Cambridge University Press.

Merkel, W. (2014). Is there a crisis of democracy? *Democratic Theory 1*(2), 11–25.

Møller, J. and S.-E. Skaaning (2013). Regime types and democratic sequencing. *Journal of Democracy 24*(1), 142–155.

Reuter, O. J. and D. Szakonyi (2019). Electoral manipulation and regime support: Survey evidence from russia. *Working paper*.

Sartori, G. (2005[1976]). *Parties and party systems: a framework for analysis.* Colchester: ECPR.

Shapiro, R. Y. and Y. Bloch-Elkon (2008). Do the facts speak for themselves? partisan disagreement as a challenge to democratic competence. *Critical Review 20*(1-2), 115–139.

Stefes, C. H. and J. Sehring (2011). Wilted roses and tulips: The regression of democratic rule in Kyrgyzstan and Georgia. In G. Erdmann and M. Kneuer (Eds.), *Regression of democracy?* Springer Science and Business Media.

Svolik, M. W. (2019). Polarization versus democracy. *Journal of Democracy 30*(3), 20–32.

Svolik, M. W. (2020). When polarization trumps civic virtue: Partisan conflict and the subversion of democracy by incumbents. *Quarterly Journal of Political Science 15*(1), 3–31.

Touchton, M., C. Klofstad, and J. Uscinski (2020). Does partisanship promote anti-democratic impulses? evidence from a survey experiment. *Journal of Elections, Public Opinion and Parties 0*(0), 1–13.

Tworzecki, H. and R. Markowski (2014). Knowledge and partisan bias: An uneasy relationship. *East European Politics and Societies and Cultures 28*(4), 836–862.

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies 69*, 102199.

Appendix for "Do Partisanship and Policy Agreement Make Citizens Tolerate Undemocratic Behavior?"

Kristian Vrede Skaaning Frederiksen*

October 19, 2021

Appendix A: Candidate Attributes, Scenario Example, and Ethical Practices

This appendix provides all candidate attributes and their probabilities of being assigned (Tables A1-A3) and a candidate choice scenario example (Figure A1). As described in the article, age, gender, and background are assigned according to the target population of the candidates whereas the attributes of theoretical interest are assigned using uniform distributions (where the probability of each attribute level is equal to one divided by the number of possible levels). Table A4 shows the statements used to measure the respondents' policy preferences. As mentioned in the article, they were asked whether they agreed or disagreed with these statements on a five-point scale. In order to conduct the policy distance and intensity measures described in the article, the statements correspond to the positions assigned to the candidates in Tables A1-A3.

As the candidate choice scenarios are what the experiments presented in the article are about, I also describe ethical practices here. Regarding deception, harm, and impact, it is clearly stated in the surveys that the candidates are fictitious and hypothetical. I chose to present fictitious candidates for the respondents in order not to let impressions about real-world candidates bias the results and, moreover, in order not to deceive respondents about political candidates that actually exist. Moreover, it is stated clearly in the survey that it is part of a research study about political attitudes. Lucid obtained consent and compensated each respondent economically for participating. Confidentially was ensured and a data processing agreement was entered into to ensure protection of personal data and to comply with GDPR law.

*PhD Candidate, Department of Political Science, Aarhus University

Table A1: Distribution of Attributes: US, UK, and CZ. Age is drawn randomly from probability-specified intervals.

Attribute	**Values (Probability)**
Age	**US**: 40-49 (0.2), 50-59 (0.2), 60-62 (0.2), 63-66 (0.2), 67-75 (0.2), **CZ**: 43-57 (0.333), 58-67 (0.333), 68-77 (0.333), **UK**: 44-53 (0.333), 54-57 (0.333), 58-61 (0.333)
Gender	**US**: Female (0.2), Male (0.8), **CZ**: Female (0.083), Male (0.917), **UK**: Female (0.286), Male (0.714)
Background	**US**: Company Director/Founder (0.4), Civil Servant (0.1), Self-employed (0.1), Lawyer (0.3), Political Career (0.1), **CZ**: Political Career (0.083), Actor (0.167), Journalist (0.25), Academic (0.5), **UK**: Lawyer (0.222), Civil Servant (0.111), Banker (0.222), Journalist (0.222), Political Career (0.222)
Party	**US**: Democrat (0.5), Republican (0.5), **CZ**: ČSSD (0.333), ODS (0.333), ANO 2011 (0.333), **UK**: Conservatives (0.5), Labour (0.5)
Positions (one per issue)	*Redistribution:* Increase/Decrease public welfare spending (0.167/0.167) Increase/Decrease power of labor (**UK**: trade) unions (0.167/0.167) Increase/Decrease income tax on 10 percent richest (0.167/0.167) *Morality Issues:* Make it easier/harder for people of the same sex to marry each other (0.167/0.167) Make it easier/harder for women to get an abortion (0.167/0.167) Allow illegal immigrants to apply for citizenship/Increase efforts to arrest and eventually deport illegal immigrants (0.167/0.167) *Undemocratic/Democratic Behavior:* Said it is legitimate to fight political opponents in the streets if one feels provoked/Said it is unacceptable to fight political opponents in the streets even though one feels provoked (0.125/0.125) Supported a proposal to reduce polling stations in areas that support opposing parties/Supported a proposal to preserve existing polling-stations in all areas (0.125/0.125) Said court rulings by judges appointed by opposing parties should be ignored/adhered to (0.125/0.125) Said it is acceptable to harass journalists that do not reveal sources/Said it is unacceptable to harass journalists even though they do not reveal sources (0.125/0.125)
Reputation (one per issue)	*Economy:* Good/Bad at handling economic matters (0.333/0.333), Neither good nor bad reputation on economic matters (0.333) *Corruption:* Bad/Good at fighting corruption (0.333/0.333), Neither good nor bad reputation on fighting corruption (0.333)

Table A2: Distribution of Attributes: Mexico

Attribute	**Values (Probability)**
Age	39 (0.031), 40 (0.031), 44 (0.063), 45 (0.031), 46 (0.031), 47 (0.031), 48 (0.031), 49 (0.063), 50 (0.094), 51 (0.031), 52 (0.094), 53 (0.031), 54 (0.031), 55 (0.031), 56 (0.063), 57 (0.094), 58 (0.031), 60 (0.031), 61 (0.063), 62 (0.031), 69 (0.062)
Gender	Female (0.062), Male (0.938)
Background	Accountant (0.125), Business Administration (0.062), Civil Servant (0.094), Engineer (0.125), Self-employed (0.094), Journalist (0.031), Lawyer (0.406), Academic (0.031), Professional Sports (0.031)
Party	MORENA (0.25), PAN (0.25), PRD (0.25), PRI (0.25)
Positions (one per issue)	*Redistribution:* Increase/Decrease public welfare spending (0.167/0.167) Provide/Prevent universal access to public colleges (0.167/0.167) Increase/Decrease income tax on 10 percent richest (0.167/0.167) *Morality Issues:* Legalize/Prohibit same-sex marriage nationally (0.167/0.167) Relax abortion law/Make abortion law more strict (0.167/0.167) Provide amnesty to low-level drug offenders/Punish all drug-related crime harsher (0.167/0.167) *Undemocratic/Democratic Behavior:* Said it is legitimate to fight political opponents in the streets if one feels provoked/Said it is unacceptable to fight political opponents in the streets even though one feels provoked (0.125/0.125) Supported a proposal to reduce polling stations in areas that support opposing parties/Supported a proposal to preserve existing polling-stations in all areas (0.125/0.125) Said court rulings by judges appointed by opposing parties should be ignored/adhered to (0.125/0.125) Said it is acceptable to harass journalists that do not reveal sources/Said it is unacceptable to harass journalists even though they do not reveal sources (0.125/0.125)
Reputation (one per issue)	*Economy:* Good/Bad at handling economic matters (0.333/0.333), Neither good nor bad reputation on economic matters (0.333) *Corruption:* Bad/Good at fighting corruption (0.333/0.333), Neither good nor bad reputation on fighting corruption (0.333)

Table A3: Distribution of Attributes: South Korea

Attributes	**Values (Probability)**
Age	49 (0.056), 54 (0.056), 55 (0.111), 56 (0.056), 59 (0.056), 60 (0.111), 62 (0.056), 63 (0.056), 64 (0.056), 65 (0.056), 66 (0.056), 67 (0.056), 68 (0.056), 72 (0.056), 73 (0.056), 77 (0.056)
Gender	Female (0.056), Male (0.944)
Background	Army General (0.056), Civil Servant (0.056), Company Director (0.056), Engineer (0.056), Self-employed (0.167), Journalist (0.056), Lawyer (0.389), Professor (0.056), Political career (0.111)
Party	UFP (0.50), DPK (0.50)
Positions (one per issue)	*Redistribution:* Increase/Decrease public welfare spending (0.167/0.167) Increase/Decrease power of labor unions (0.167/0.167) Increase/Decrease income tax on 10 percent richest (0.167/0.167) *Morality Issues:* Legalize/Prohibit same-sex marriage nationally (0.167/0.167) Relax abortion law/Make abortion law more strict (0.167/0.167) Increase/Decrease funds to the army (0.167/0.167) *Undemocratic/Democratic Behavior:* Said it is legitimate to fight political opponents in the streets if one feels provoked/Said it is unacceptable to fight political opponents in the streets even though one feels provoked (0.125/0.125) Supported a proposal to reduce polling stations in areas that support opposing parties/Supported a proposal to preserve existing polling-stations in all areas (0.125/0.125) Said court rulings by judges appointed by opposing parties should be ignored/adhered to (0.125/0.125) Said it is acceptable to harass journalists that do not reveal sources/Said it is unacceptable to harass journalists even though they do not reveal sources (0.125/0.125)
Reputation (one per issue)	*Economy:* Good/Bad at handling economic matters (0.333/0.333), Neither good nor bad reputation on economic matters (0.333) *Corruption:* Bad/Good at fighting corruption (0.333/0.333), Neither good nor bad reputation on fighting corruption (0.333)

Table A4: Statements for measurement of respondents' policy preferences (three statements per dimension per country)

Redistribution:
Taxes on the richest people in this country should be decreased (US, UK, MX, CZ, SK)
Public welfare spending should be increased (US, UK, MX, CZ, SK)
Labor (UK: Trade) unions have too little power (US, UK, CZ, SK)
The government should provide universal access to public colleges (MX)
Morality Issues:
It should be easier for women to get an abortion (US, UK, CZ)
Abortion should always be legal (MX, SK)
It should be harder for people of the same sex to marry each other (US, UK, CZ)
Same-sex marriage should be prohibited (MX, SK)
Illegal immigrants should be allowed to apply for U.S./British/Czech citizenship (US, UK, CZ)
The government should provide amnesty to low-level drug offenders (MX)
Government funds to the army should be decreased (SK)

Figure A1: Candidate choice example: US.

Imagine that the following two candidates run for the 2024 presidential election. Please read the candidate profiles carefully, and tell us how likely you would be to vote for each of the two candidates at the election.

	Candidate 1	**Candidate 2**
Age	66	44
Gender	Male	Male
Background	Lawyer	Company Founder/Director
Party	Republican	Democrat
Positions	- Decrease public welfare spending - Make it harder for women to get an abortion - Supported a proposal to reduce polling stations in areas that support opposing parties	- Decrease public welfare spending - Make it easier for women to get an abortion - Said it is legitimate to fight political opponents in the streets if one feels provoked
Reputation	- Good at handling economic matters - Good at fighting corruption	- Bad at handling economic matters - Good at fighting corruption

	How likely is it that you would vote for **candidate 1**?	How likely is it that you would vote for **candidate 2**?
Very likely	○	○
Somewhat likely	○	○
Neither likely nor unlikely	○	○
Somewhat unlikely	○	○
Very unlikely	○	○
Don't know	○	○

→

Appendix B: Supplementary Tables and Figures for Main Findings

This appendix provides regression tables as an alternative presentation of the results in Figures 1-3 and elaborated graphic illustrations of the levels of support for undemocratic and democratic candidates examined in Section 4.2 of the article.

Tables B1-B3 show the interactions between undemocratic behavior and partisanship, policy distance, and preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate. Figures B1-B3 show the levels of support for undemocratic and democratic candidates across the partisanship and policy agreement measures. The model specifications are the exact same as in Figures 1-3 and Table 3.

Appendix C: Robustness to Splitting the Undemocratic Behavior, Policy Distance, and Preference Intensity Measures

Expanding on the original Figures 1-3, this appendix provides more nuanced and complex analyses splitting the policy distance and preference intensity measures by their two dimensions — morality issues and redistribution — and splitting the undemocratic behavior measure by its four antagonistic pairs. The wordings of the individual attributes of these measures are shown in Table 1 in the article and in Tables A1-A3 in Appendix A. To keep the number of graphs somewhat down, I only plot the pooled estimates when splitting the undemocratic behavior measure.

In sum, I show that the main findings from Figures 1-3 are robust to the splits. Moreover, I utilize the differing effect sizes of the individual undemocratic behavior-items to illustrate the sensitivity of the finding that voters prefer politically favorable undemocratic candidates over politically unfavorable democratic candidates.

Table B1: Effect of undemocratic behavior on support for incumbent candidates conditioned by partisanship in the Czech Republic, Mexico, South Korea, the United Kingdom, and the United States.

	CZ	MX	SK	UK	US	Pooled
Undemocratic Behavior	-0.14***	-0.09***	-0.09**	-0.15***	-0.24***	-0.14***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.01)
Partisanship (Ref.: Dislike a great deal)	0.00	0.00	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)	(.)	(.)
Dislike somewhat	0.50***	0.59***	0.52***	0.45***	0.43***	0.50***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.02)
Neutral	0.69***	0.84***	0.76***	0.73***	0.68***	0.74***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)
Like somewhat	0.86***	1.05***	0.98***	0.89***	0.93***	0.95***
	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.02)
Like a great deal	1.25***	1.59***	1.44***	1.22***	1.35***	1.42***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.04)	(0.02)
Undemocratic x Dislike somewhat	-0.08	-0.08*	-0.00	-0.07	-0.01	-0.05**
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.02)
Undemocratic x Neutral	-0.00	-0.07*	0.03	0.01	0.12**	0.02
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)
Undemocratic x Like somewhat	-0.08*	-0.06	0.03	-0.04	-0.00	-0.04*
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.02)
Undemocratic x Like a great deal	0.06	-0.03	0.03	0.01	0.06	-0.00
	(0.06)	(0.05)	(0.05)	(0.04)	(0.03)	(0.02)
Constant	1.98***	2.11***	2.11***	2.01***	2.29***	2.10***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.01)
Adjusted R^2	0.098	0.122	0.101	0.097	0.128	0.119
n (Candidates)	41,788	51,081	43,349	49,721	53,417	239,356
N (Respondents)	2,350	2,784	2,417	2,758	2,999	13,308

Unstandardized regression coefficients.

Respondent clustered standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B2: Effect of undemocratic behavior on support for incumbent candidates conditioned by policy distance in the Czech Republic, Mexico, South Korea, the United Kingdom, and the United States.

	CZ	MX	SK	UK	US	Pooled
Undemocratic Behavior	-0.23***	-0.24***	-0.15*	-0.30***	-0.38***	-0.27***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.02)
Policy Distance	-0.22***	-0.18***	-0.14***	-0.29***	-0.30***	-0.24***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)
Undemocratic x Distance	0.01	0.03	0.05	0.06*	0.08**	0.04**
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.01)
Policy Distance^2	0.01*	0.00	0.01	0.01***	0.02***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Undemocratic x Distance^2	0.00	-0.00	-0.01	-0.00	-0.01	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	3.11***	3.34***	3.17***	3.43***	3.75***	3.40***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.02)
Adjusted R^2	0.039	0.032	0.010	0.054	0.030	0.032
n (Candidates)	41,788	51,081	43,349	49,721	53,417	239,356
N (Respondents)	2,350	2,784	2,417	2,758	2,999	13,308

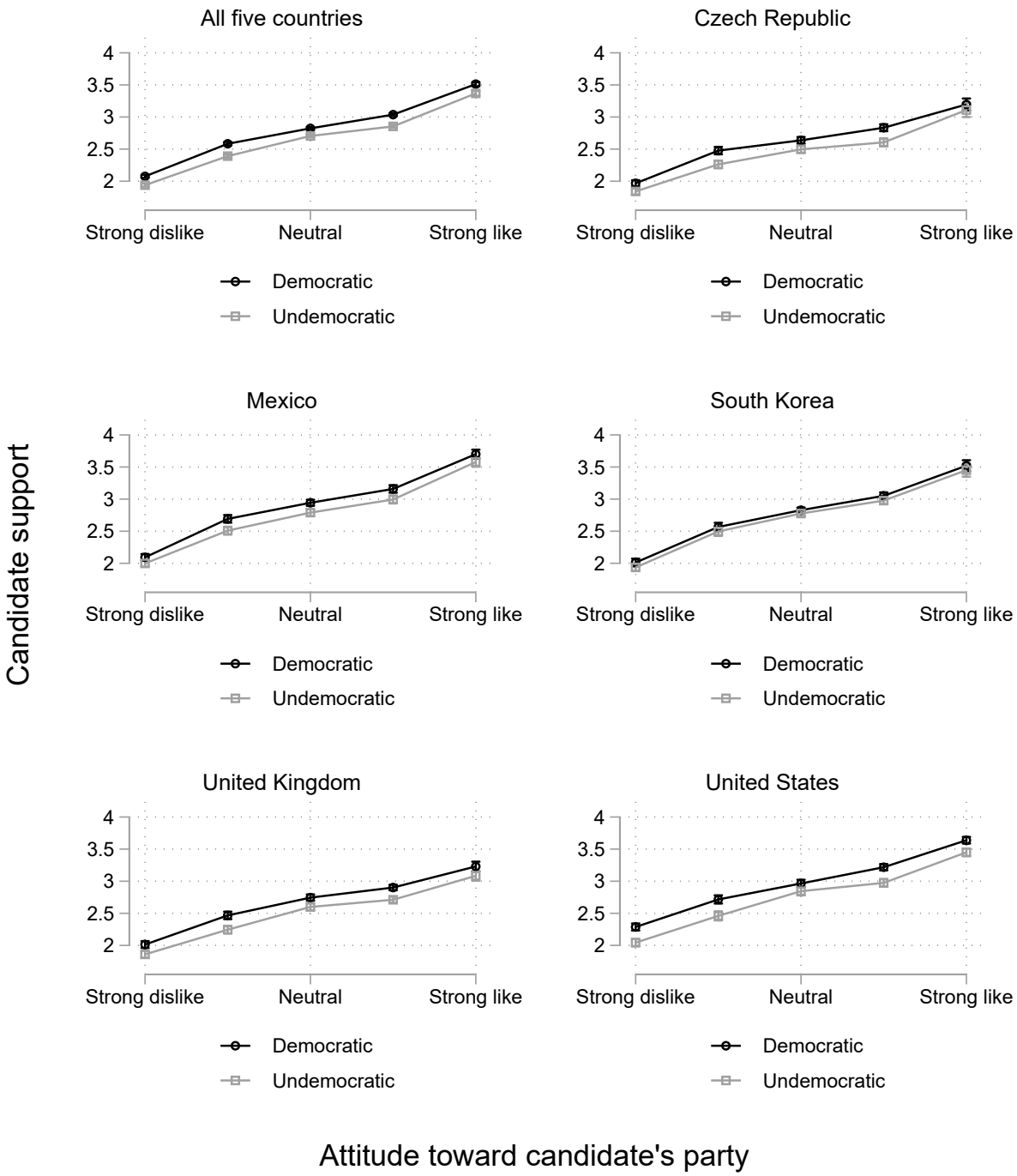
Unstandardized regression coefficients.
Respondent clustered standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B3: Effect of undemocratic behavior on support for incumbent candidates conditioned by preference intensity in the Czech Republic, Mexico, South Korea, the United Kingdom, and the United States.

	CZ	MX	SK	UK	US	Pooled
Undemocratic Behavior	-0.18***	-0.15***	-0.06***	-0.17***	-0.21***	-0.15***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Intensity	0.11***	0.08***	0.06***	0.12***	0.10***	0.09***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Undemocratic x Intensity	-0.01*	-0.01**	-0.00	-0.01**	-0.02***	-0.01***
	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Intensity^2	0.00	0.00	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Undemocratic x Intensity^2	0.00	0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	2.49***	2.76***	2.70***	2.65***	2.99***	2.72***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Adjusted R^2	0.032	0.018	0.012	0.038	0.029	0.025
n (Candidates)	41,788	51,081	43,349	49,721	53,417	239,356
N (Respondents)	2,350	2,784	2,417	2,758	2,999	13,308

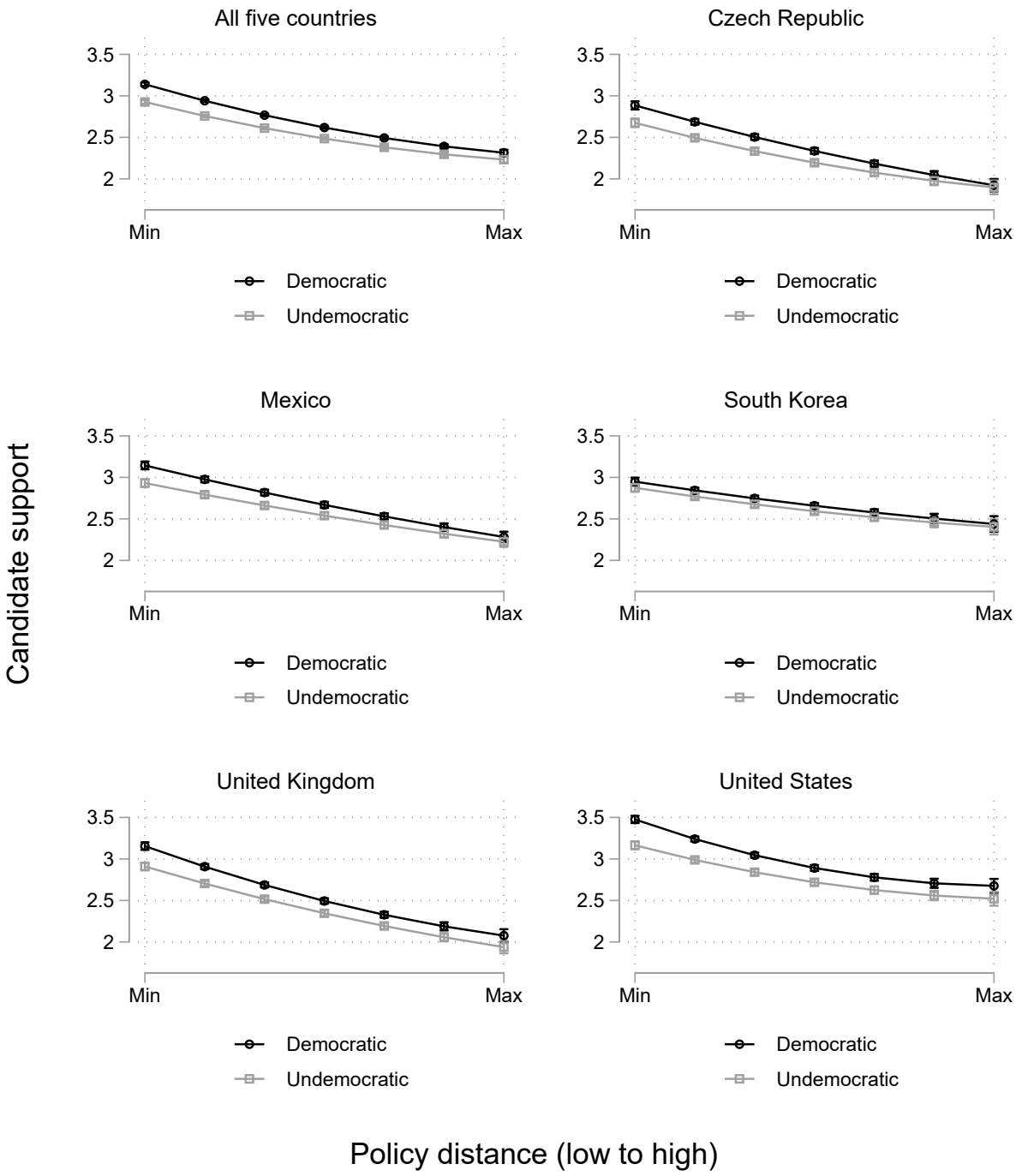
Unstandardized regression coefficients.
Respondent clustered standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure B1: Levels of support for undemocratic and democratic candidates across shared partisanship between voter and candidate.



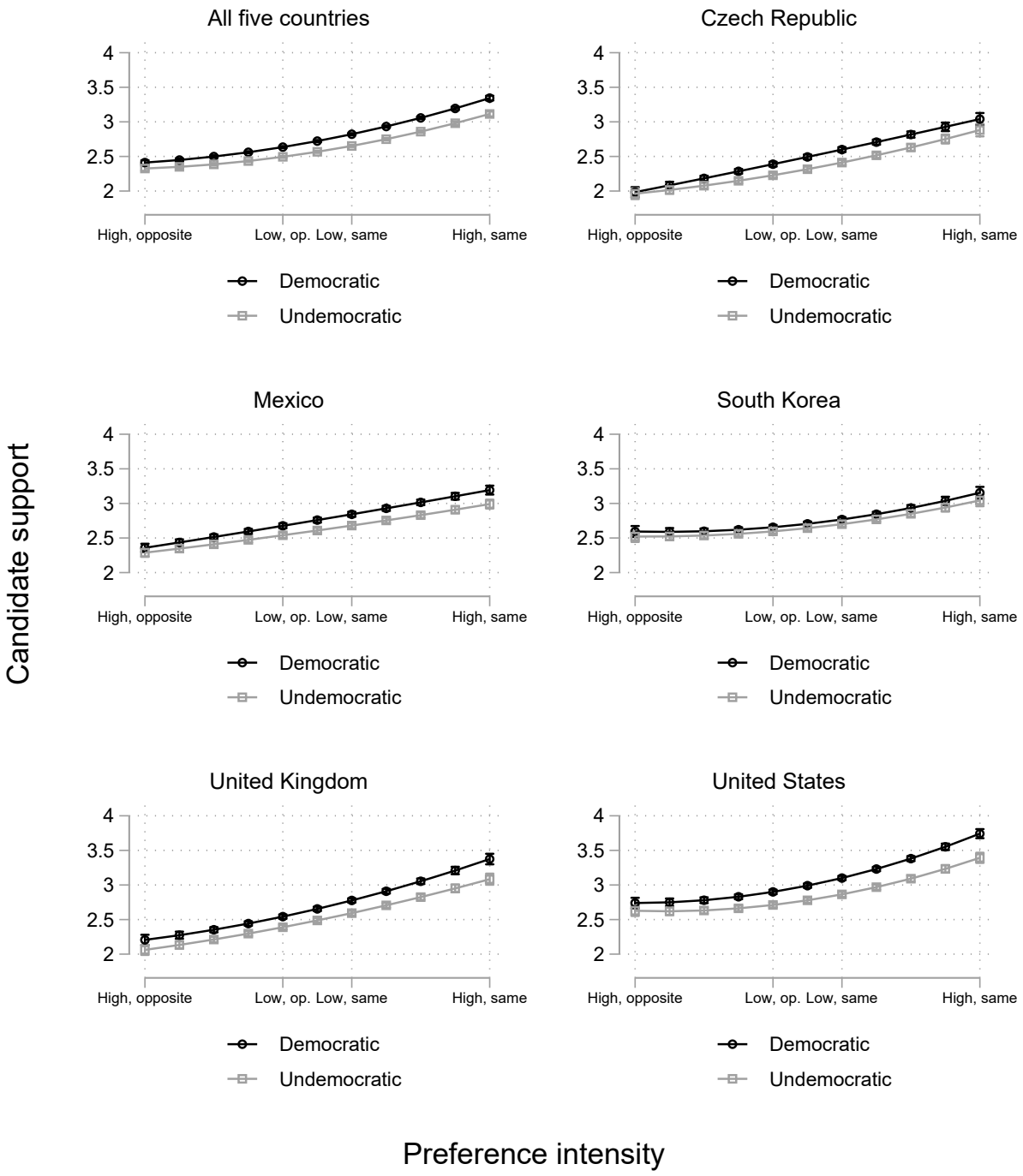
Note: Same sample and model specifications as Figure 1.

Figure B2: Levels of support for undemocratic and democratic candidates across policy distance between voter and candidate.



Note: Same sample and model specifications as Figure 2.

Figure B3: Levels of support for undemocratic and democratic candidates across preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate.



Note: Same sample and model specifications as Figure 3.

Splitting the Policy Distance Measure

In Figure C1, I show the effects of undemocratic behavior across policy distance between voter and candidate where distance is split into its two dimensions of morality issues and redistribution. The figure is, therefore, comparable to the original Figure 2. We see that the finding that voters sanction policy proximate candidates the most for behaving undemocratically holds across the two dimensions We also see, however, that this tendency seems to be most pronounced on policy distance on morality issues.

Splitting the Preference Intensity Measure

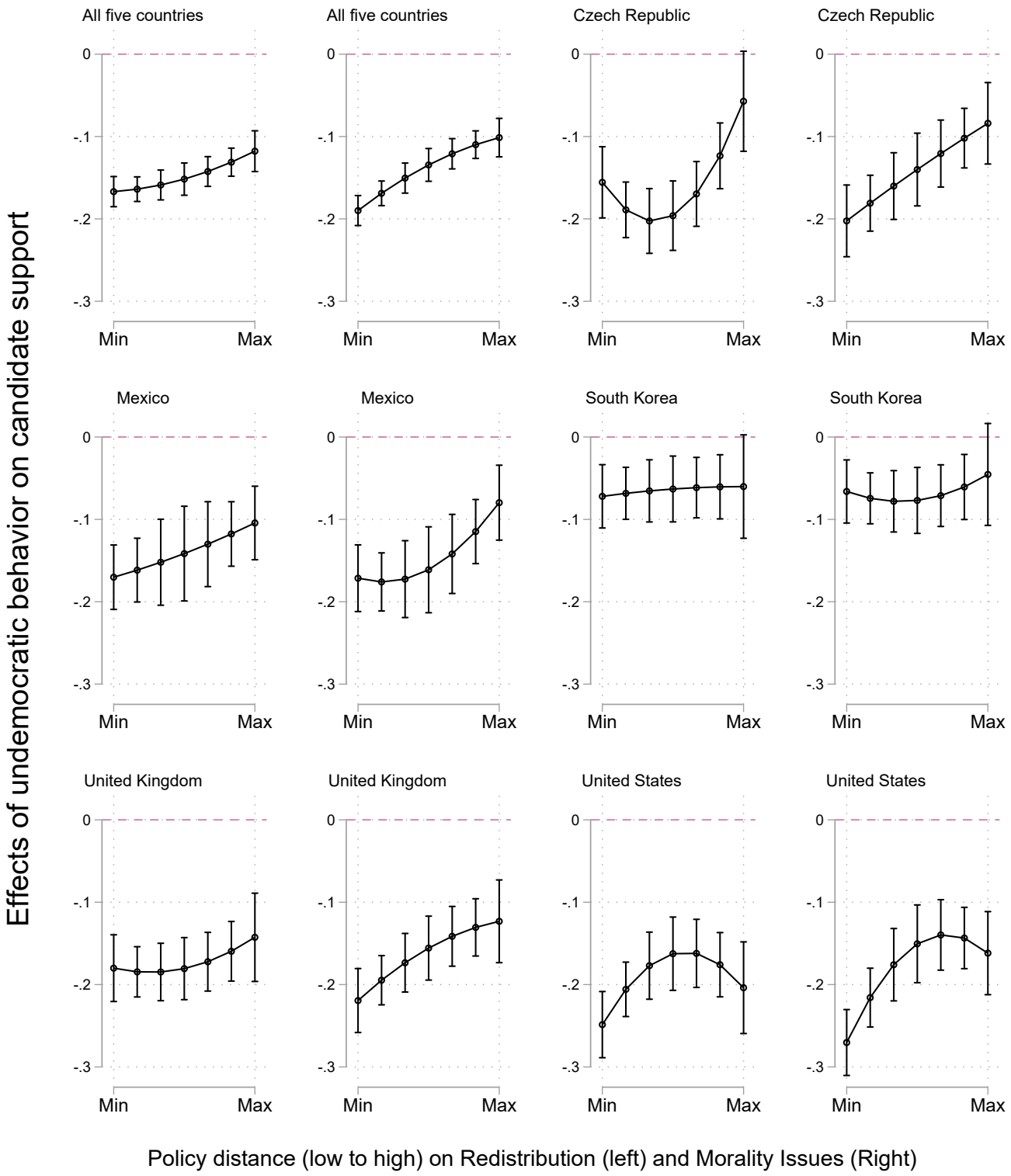
Analogously to in Figure C1, I split the preference intensity measure by its two dimensions and plot the effects of undemocratic behavior in Figure C2, which therefore is comparable to the original Figure 3. And we can essentially draw the same inferences as we did from Figure C1. First, the effect of undemocratic behavior increases with the extent to which the voter has intense preferences in favor of the candidate on both dimensions. Second, this tendency is most pronounced on morality issues.

Splitting the Undemocratic Behavior Measure

In Figures C3A-C3D, I show the effects of undemocratic behavior across partisanship, policy distance, and preference intensity when undemocratic behavior is split into its four antagonistic pairs. As shown in Table 1 in the article and in Tables A1-A3 in Appendix A, the four types of undemocratic behavior are supporting electoral manipulation, discarding judges appointed by opposing parties, encouraging violence, and endorsing journalist harassment.

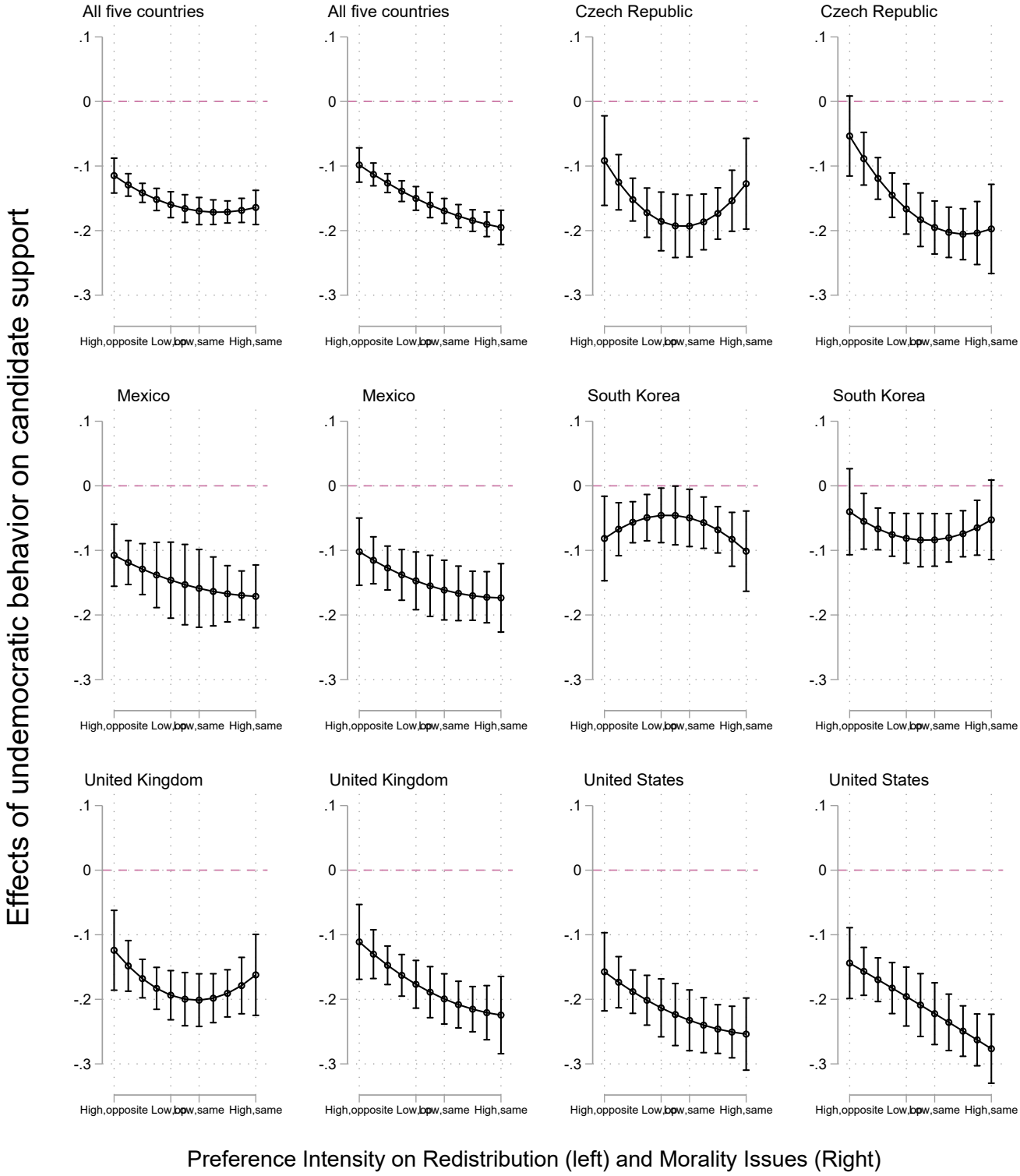
We see that the tendencies that voters do not sanction undemocratic behavior less when they hold strong feelings toward the candidate's party and that they sanction policy-proximate candidates the most for undemocratic behavior for the most part replicate across the four pairs. We see some signs of a partisan bias on the electoral manipulation item as the effect of undemocratic behavior is slightly smaller for strongly

Figure C1: Effects of undemocratic behavior on candidate support across policy distance split in redistribution and morality issues.



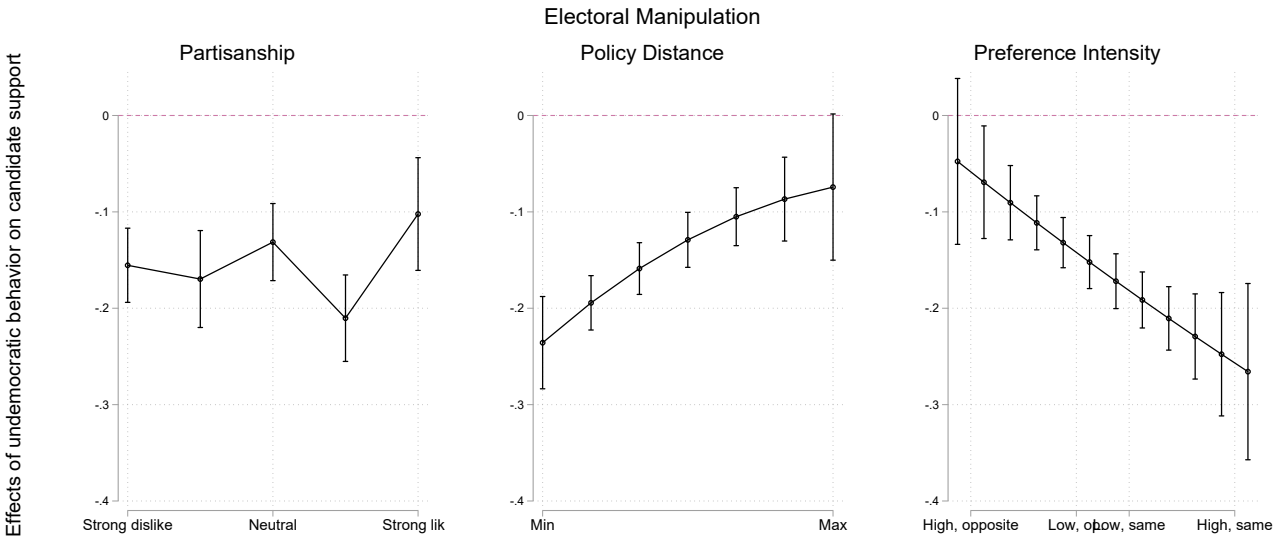
Note: OLS-regressions including interaction terms between undemocratic behavior and the two respective distance measures employing the exact same sample as in Figures 1-3 in the article.

Figure C2: Effects of undemocratic behavior on candidate support across preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate. Intensity measure split in redistribution and morality issues.



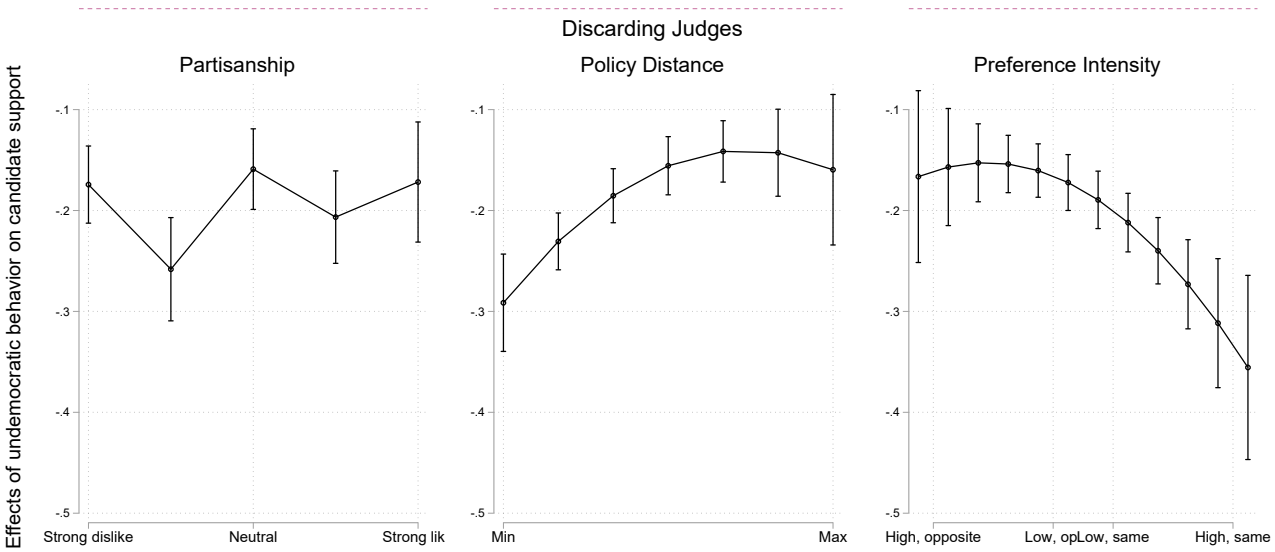
Note: OLS-regressions including interaction terms between undemocratic behavior and the two respective intensity measures employing the exact same sample as in Figures 1-3 in the article.

Figure C3A: Effects of undemocratic behavior as supporting electoral manipulation across shared partisanship between voter and candidate, policy distance between voter and candidate, and preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate.



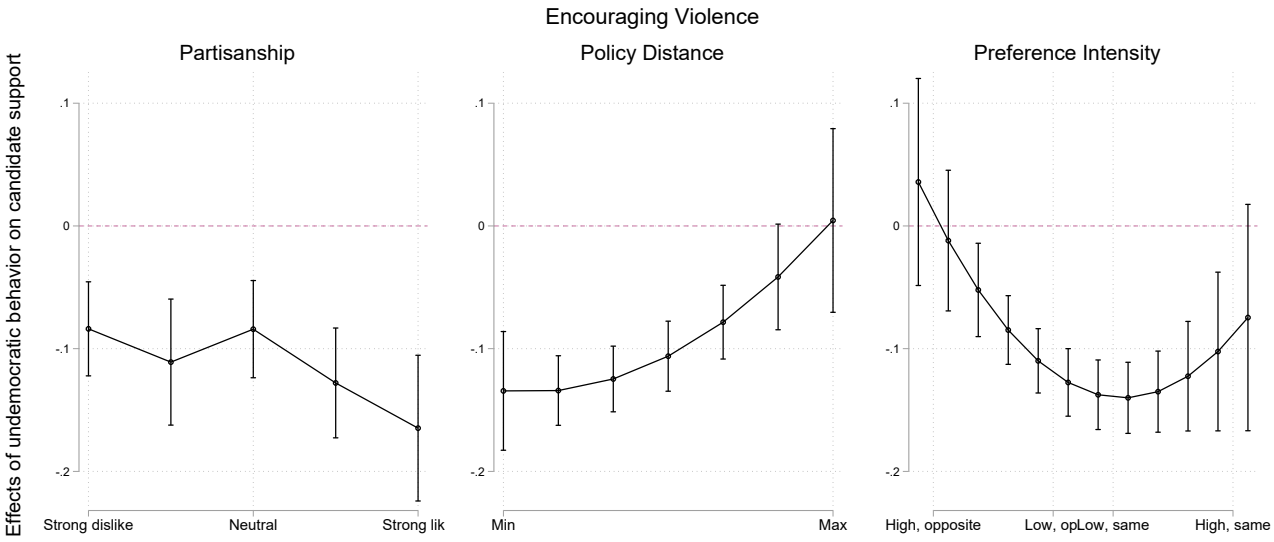
Note: OLS-regression including interaction terms between the electoral manipulation-variable and partisanship, policy distance, and preference intensity. The sample is, therefore, around one fourth the size of the original sample employed in the original Figures 1-3. All five countries included.

Figure C3B: Effects of undemocratic behavior as discarding opposing judges across shared partisanship between voter and candidate, policy distance between voter and candidate, and preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate.



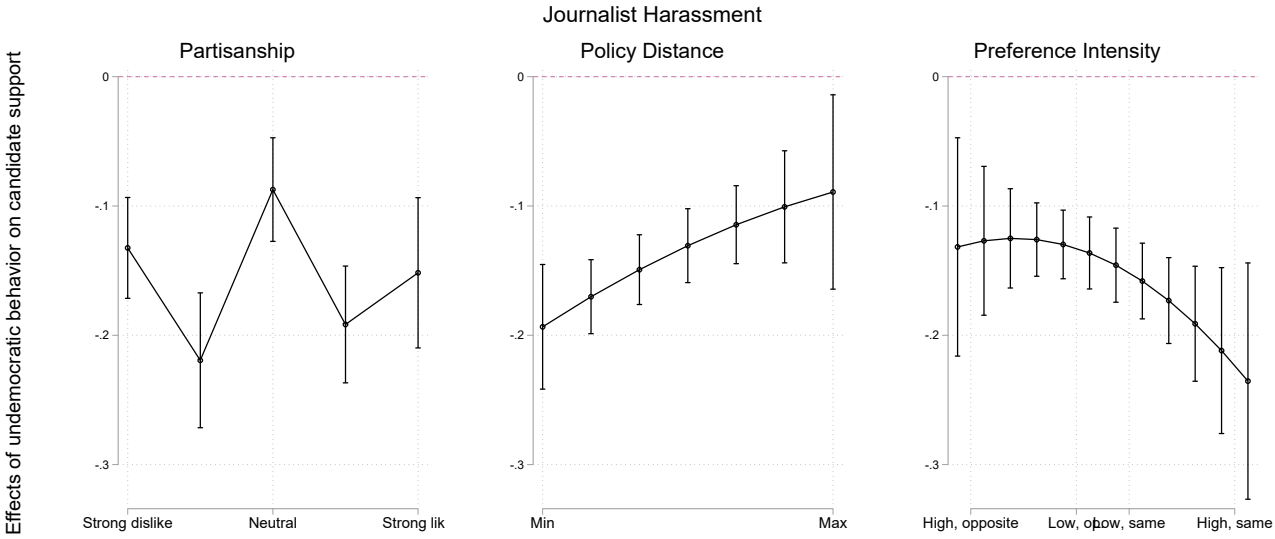
Note: OLS-regression including an interaction term between the discarding judges-variable and partisanship, policy distance, and preference intensity. The sample is, therefore, around one fourth the size of the original sample employed in the original Figures 1-3. All five countries included.

Figure C3C: Effects of undemocratic behavior as encouraging violence across shared partisanship between voter and candidate, policy distance between voter and candidate, and preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate.



Note: OLS-regression including an interaction term between the encourage violence-variable and partisanship, policy distance, and preference intensity. The sample is, therefore, around one fourth the size of the original sample employed in the original Figures 1-3. All five countries included.

Figure C3D: Effects of undemocratic behavior as legitimizing journalist harassment across shared partisanship between voter and candidate, policy distance between voter and candidate, and preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate.



Note: OLS-regression including an interaction term between the journalist harassment-variable and partisanship, policy distance, and preference intensity. The sample is, therefore, around one fourth the size of the original sample employed in the original Figures 1-3. All five countries included.

in-partisan candidates there, but on the same item the impact is largest when the voter has moderately positive feelings toward the candidate's party. Meanwhile, on encouraging violence, the effect of undemocratic behavior is largest for strongly in-partisan candidates.

Finally, we see that the encouraging violence item generally produces the smallest effects while the discarding judges item produces the largest effects. This difference provides an opportunity to explore the consequences of the strength of experimental treatments for the finding that voters are willing to compromise democratic compliance for partisan ends. For example, moderately in-partisan undemocratic candidates beat democratically compliant candidates from parties that the voter has neutral feelings toward when focusing on encouraging violence (diff. = 0.06, $p < 0.01$).[1] But when focusing on discarding opposing judges, this match-up ends in a draw (diff. = -0.00, $p = 0.89$). Similarly, a democratic candidate who scores 2 on policy distance beats an undemocratic candidate scoring 1 on distance when focusing on discarding opposing judges (diff. = -0.08, $p < 0.001$) but loses narrowly when focusing on encouraging violence (diff. = 0.04, $p = 0.05$). These findings illustrate the sensitivity — to the strength of experimental treatments, that is — of the finding that voters prefer politically favorable undemocratic candidates over politically unfavorable democratic candidates.

Appendix D: Alternative Measure of Partisanship

In this appendix, I substitute the original measure of shared partisanship between voter and candidate with an alternative measure. Contrary to the original measure, which is based on affective like/dislike-scores, this measure categorically asks the respondents which party they identify with (if any) and simply groups the respondent-candidate dyads in 'Co-partisanship', 'Opposing partisanship', and 'Non-partisan'.[2] In Figure D1, I show

[1] In this section, negative differences signal that the democratic candidate wins while positive differences signal that the undemocratic candidate wins.

[2] Specifically, the respondents where asked if they identify with any of the parties in the political system of their country. If the respondents answered that they identify with a particular party that where then assigned to a particular candidate (who they only encountered afterwards), I coded that particular candidate-observation as 'Co-partisanship'. If the respondents answered that they identified with another party (or a party that was not listed as an option), I coded the observation as 'Opposing partisanship'. Finally, if the respondents answered that they do not identify with any party, I coded the observation as 'Non-partisan'.

the effects of undemocratic behavior across this alternative measure of partisanship. The figure is, therefore, directly comparable to the original Figure 1.

We generally see that the results do not change if we use this alternative measure. In Figure D1, we see that voters do not show any bias in favor of co-partisan candidates when sanctioning undemocratic behavior as the effects are not smallest among these candidates. The only remarkable finding we see is, contrarily, that American voters sanction co-partisan candidates more than opposing candidates for behaving undemocratically. This finding was not evident when employing the original affective measure.

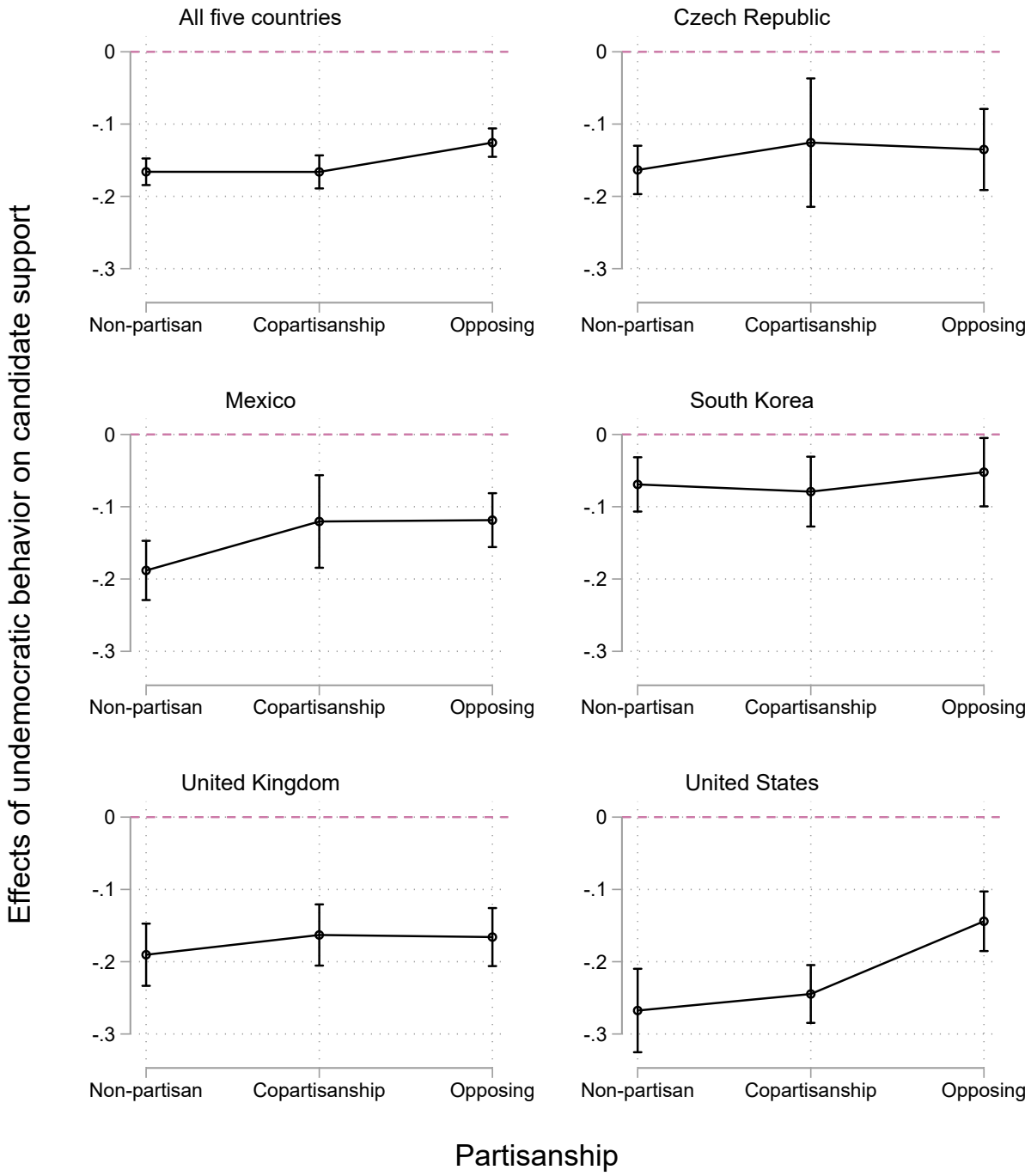
Appendix E: Adjustment of Policy Agreement Measures

This appendix includes tests with minor adjustments to the policy distance and preference intensity measures. I provide two adjusted policy distance measures where the candidate positions are assumed to be more or less extreme, respectively, than in the original measure. In the original measure, left-wing candidate positions were coded as 1.5 on the five-point scale that the respondents were placed on on the basis of answers to the statements shown in Table A4. Right-wing positions were coded as 4.5. In the two alternative measures provided in this appendix, left-wing positions are coded as 1 and 2, respectively, while right-wing positions are coded as 4 and 5, respectively.

I provide one adjusted preference intensity measure. In the original measure, observations where left-wing (right-wing) respondents encountered right-wing (left-wing) candidates were coded as -1 while same-side dyads were coded as 1 (i.e., before multiplying with the variables measuring how important the particular issues are to the respondents). If respondents took on a neutral position on a particular issue, the observation was coded as -0.5. In the alternative measure provided here, I remove these neutrals from the regression.

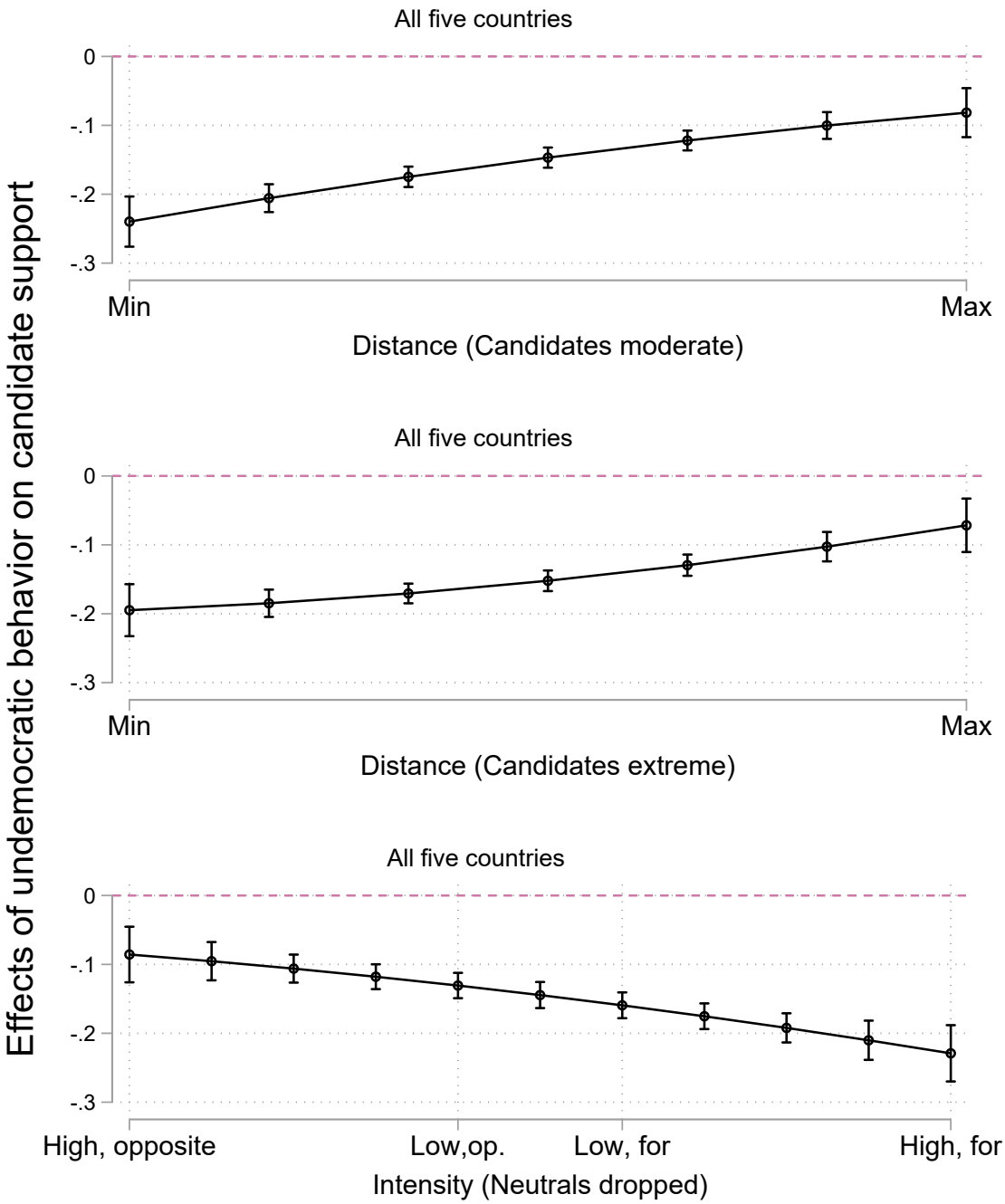
Figure E shows the results when using these three alternative measures. I only show the pooled estimates to keep the test fairly concise. Thus, the figure shows the pooled effects of undemocratic behavior conditioned by the three measures. We see the same picture as in the original specifications: Voters sanction undemocratic behavior the

Figure D1: Effects of undemocratic behavior on candidate support across shared partisanship between voter and candidate. Alternative, categorical measure of partisanship.



Note: Same model and sample as the original Figures 1 and 4 except that the affective partisanship measure is substituted with an alternative, categorical one.

Figure E: Effects of undemocratic behavior on candidate support across adjusted policy agreement measures.



Note: Same model as the original Figures 2-3 except that the original policy distance and preference intensity measures are substituted with slightly adjusted measures as described above. The sample is slightly reduced when using the intensity measure as neutral respondents are dropped.

most if they agree with the candidate on policy. The results are, therefore, robust to these adjustments of the distance and intensity measures.

Appendix F: (Non)-Parametric Estimations

In the original specifications, I employ the policy distance and preference intensity measures in squared form while I estimate the interaction between undemocratic behavior and partisanship non-parametrically. In this appendix, I provide tests turning these choices upside down: That is, I use policy distance and intensity as categorical variables (i.e., estimate the interactions non-parametrically) and use partisanship in its squared form. I simply provide the latter test for good measure as I stated in the pre-registration that I would use all three measures in squared form. As the original Figure 4 also shows, the polynomial estimation clearly does not fit the data when it comes to partisanship. Finally, as the original variable takes on too many values to treat it categorically, I cut the alternative intensity measure in seven groups (thus resembling the distance measure).

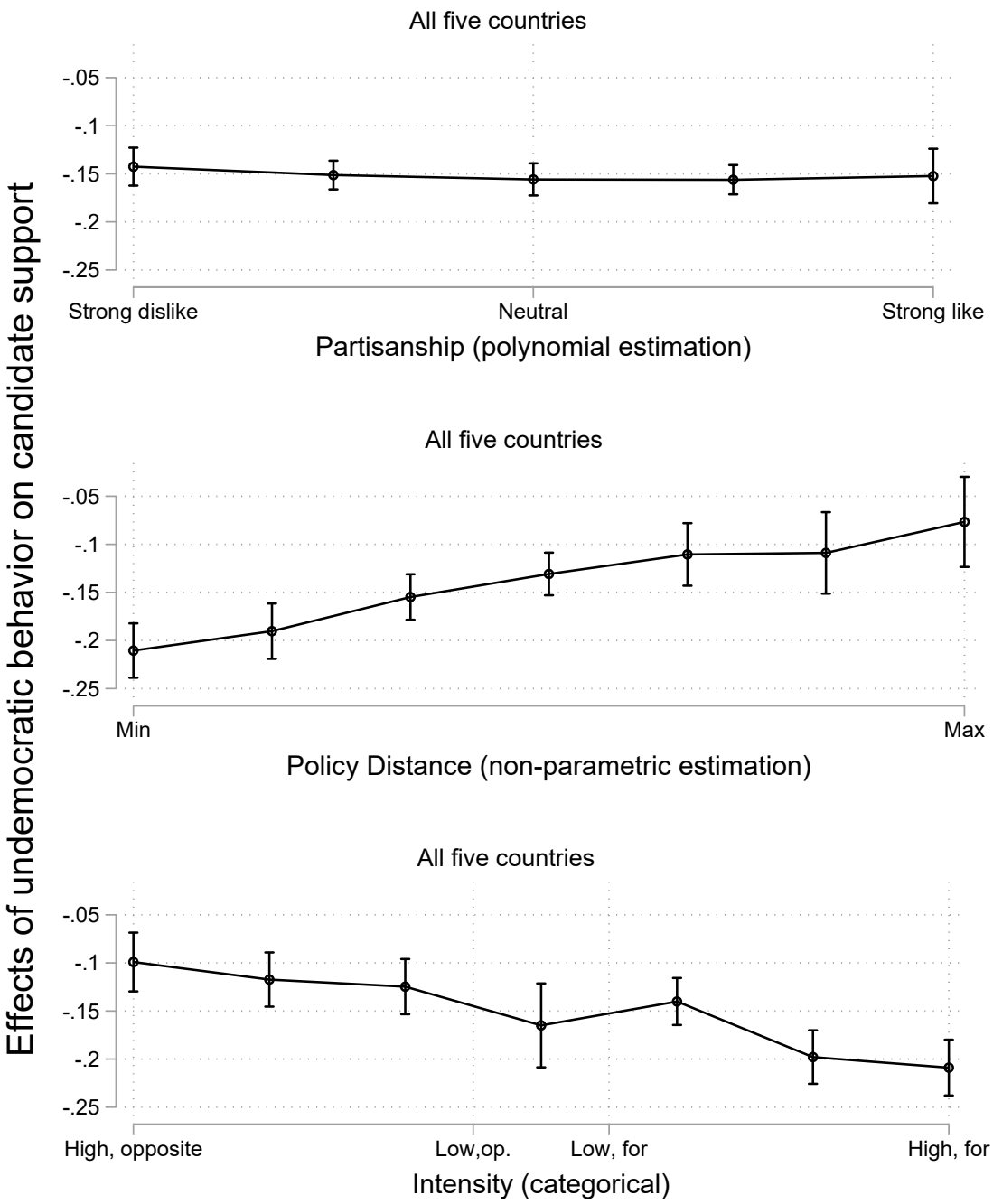
As in Appendix E, I only show the pooled estimates here due to space concerns. Thus, Figure F shows the pooled effects of undemocratic behavior the three measures. We see the same picture as in the original results: Partisanship does not diminish the negative impact of undemocratic behavior while voters sanction candidates with whom they agree on policy for behaving undemocratically the most.

Appendix G: Including Respondent Covariates

In this appendix, I include respondent covariates in the regressions which — potentially — could confound the moderating impacts of partisanship and policy agreement. The covariates included here are education (three groups corresponding to below high school, high school, and above high school), urban/rural residence (three groups equal to rural, small city, and metropolis), and gender. Beyond including average effects of these variables, I also interact them with undemocratic behavior just like I do with the partisanship and policy agreement measures.

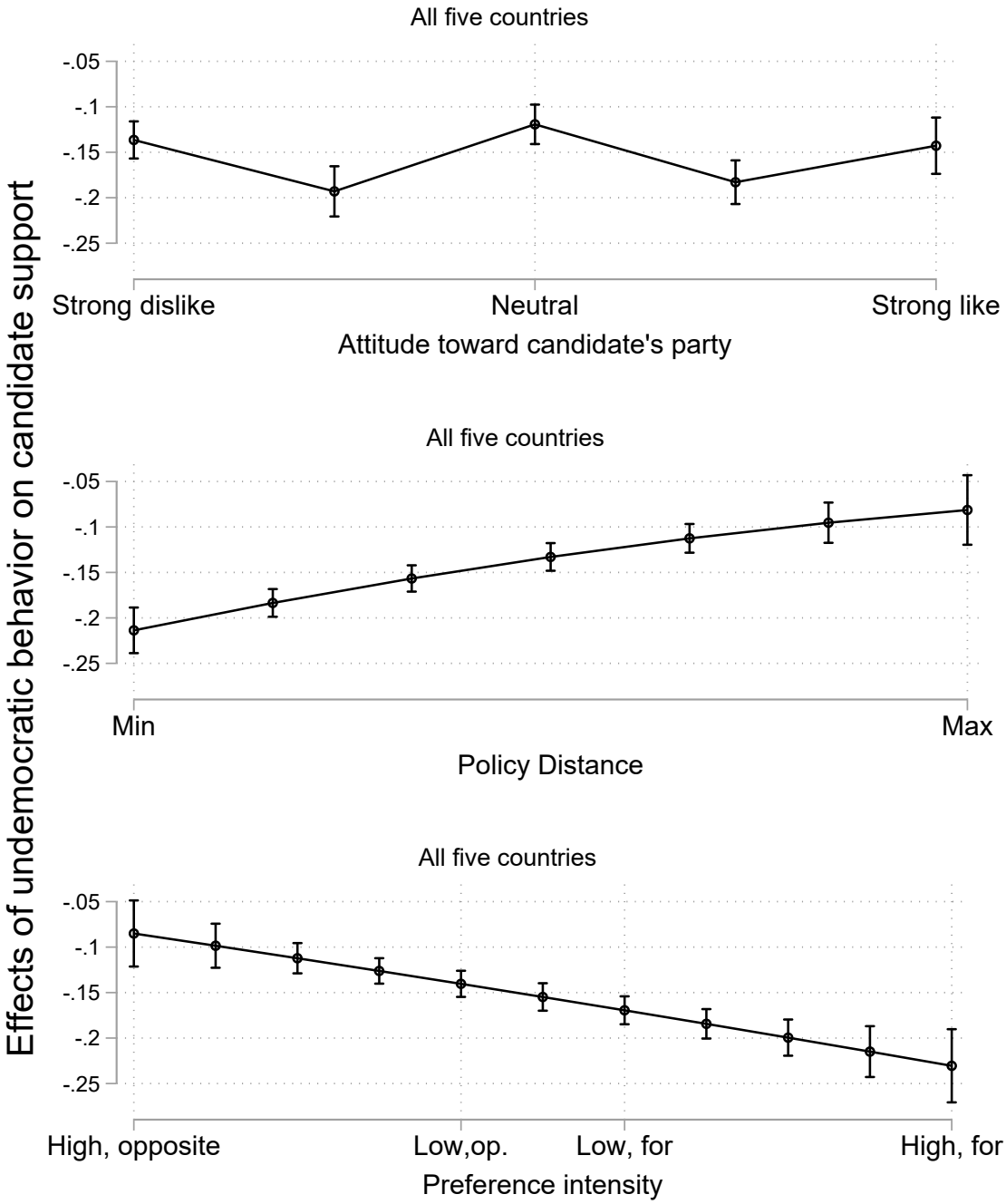
Figure G shows the pooled effects of undemocratic behavior across the partisan-

Figure F: Effects of undemocratic behavior on candidate support across partisanship, policy distance, and preference intensity. Partisanship is employed in its squared form while preference intensity and policy distance is employed as categorical variables.



Note: Same model and sample as in original Figures 1-3 except that partisanship is employed in its squared form while policy distance and preference intensity are employed categorically.

Figure G: Effects of undemocratic behavior on candidate support across partisanship, policy distance, and preference intensity. Respondent covariates (education, urban/rural residence, and gender) are included.



Note: Same model and sample as in original Figures 1-3 except that respondent covariates are included (by their average effects as well as interactions with undemocratic behavior).

ship and policy agreement measures. We see that the results are robust to this adjustment: Voters still sanction co-partisans as much as out-partisans for behaving undemocratically and even sanction candidates with whom they agree on policy the most.

Appendix H: Inclusion of Other Attributes in Specifications

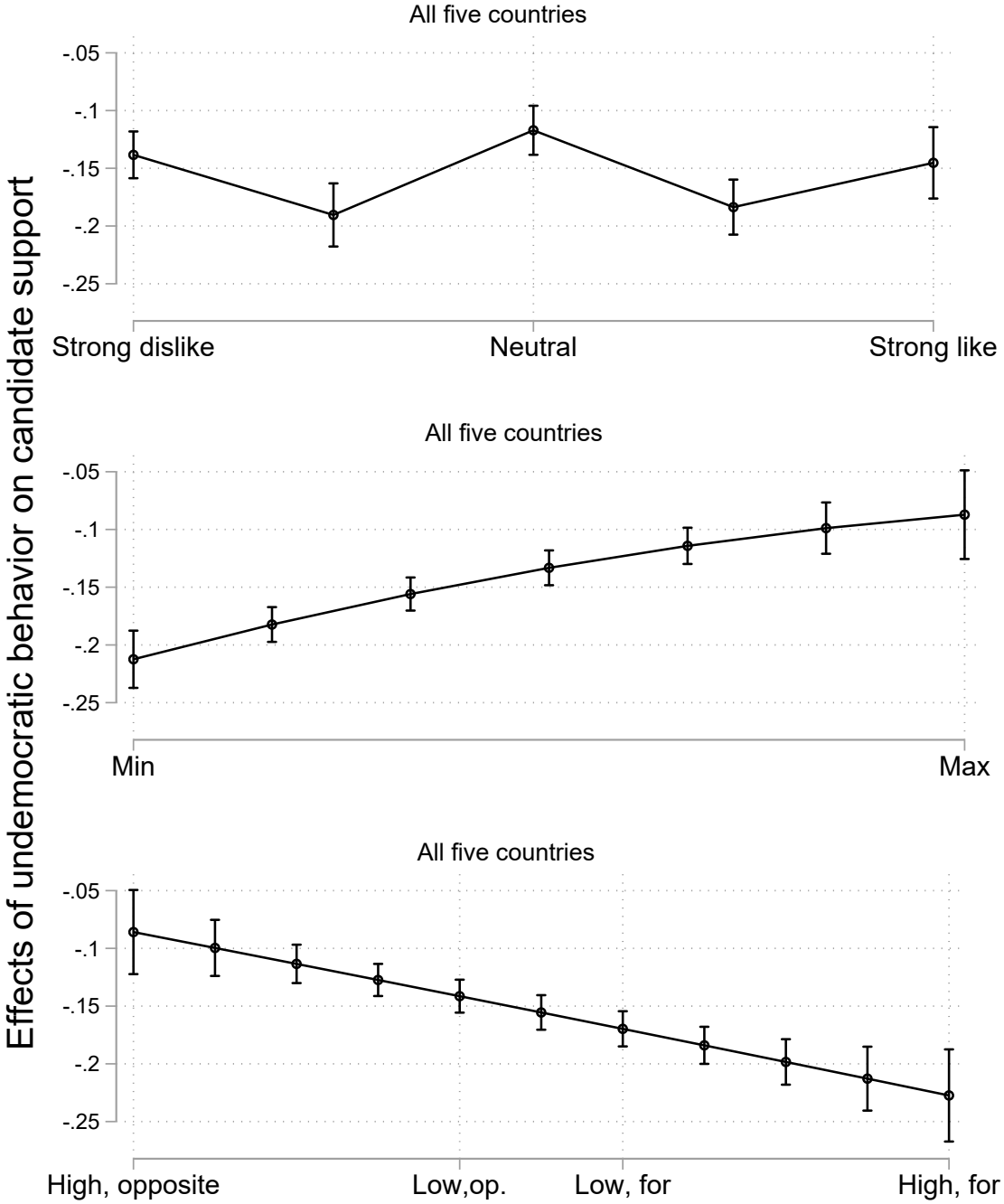
In this appendix, I add the remainder of the candidate attributes (age, gender, competence reputations, and background) to the original specifications. Thus, Figure H shows the pooled effects of undemocratic behavior across partisanship and policy agreement between voter and candidate when including these remaining attributes. We see that voters still sanction policy-proximate candidates the most while partisanship does not diminish the negative effects of undemocratic behavior.

Appendix I: Standard AMCE-plot

In this appendix, I provide a standard AMCE-plot similar to those usually shown in candidate choice studies. I have in Figure I employed a specification including undemocratic behavior, policy distance, co-partisanship, candidate age, candidate gender, and candidate competence (i.e., a scale summing competences in handling economic matters and fighting corruption) to compare the effects of these attributes. To ensure full comparability, I have rescaled all scales (distance, partisanship, intensity, and competence) to a range of 1 to 5. Candidate background is omitted as this variable does not have similar categories across countries.

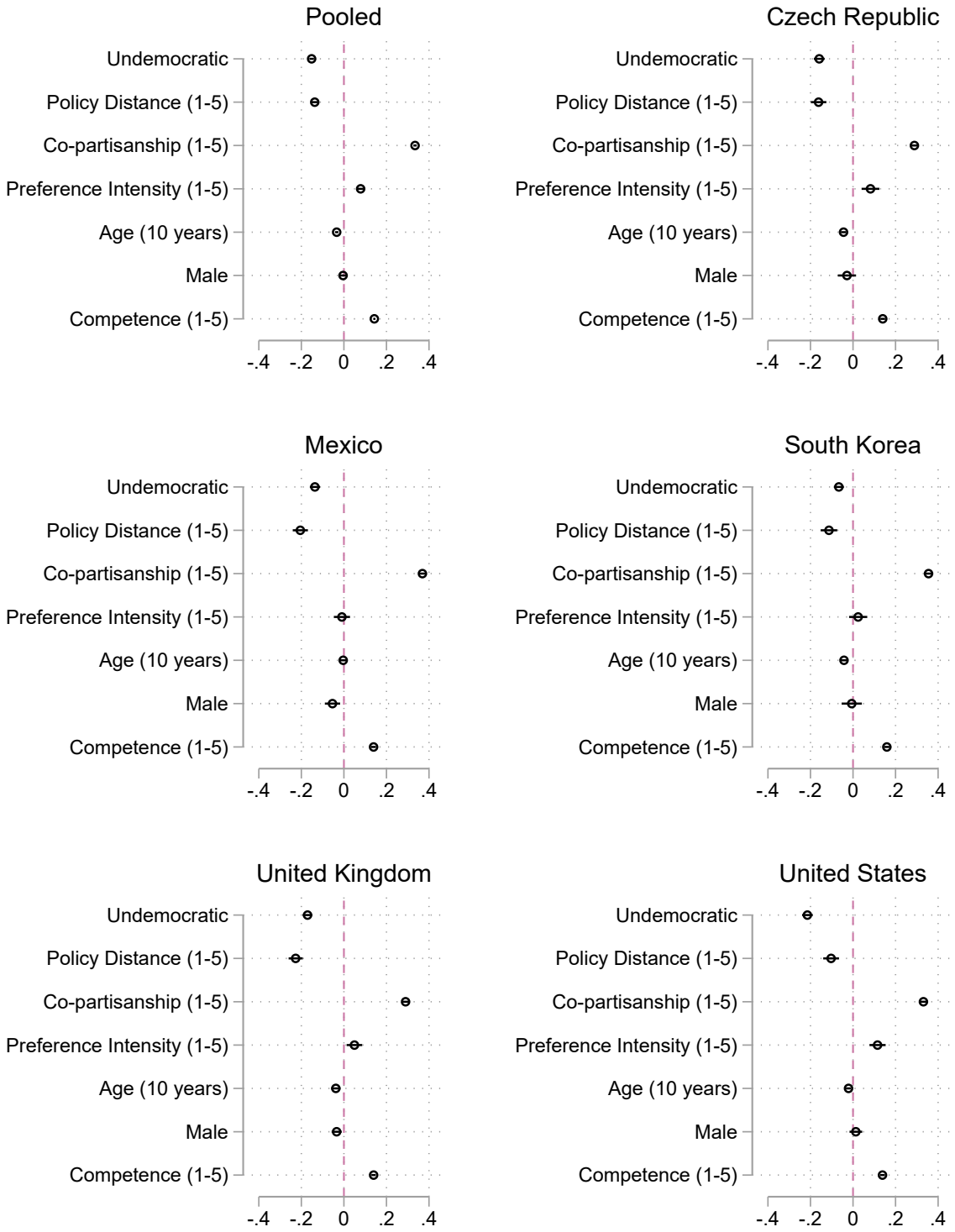
We see that the impact of one scale point of policy distance is at least as large as that of undemocratic behavior while the impact of one scale point of partisanship is markedly larger. We also see that preference intensity (i.e., to which extent the respondent holds intense preference to the same or opposite side of the candidate) produces smaller effects than policy distance — at least when both measures are included simultaneously as they are here. The effect of one scale point on the competence scale is approximately equal to that of undemocratic behavior, while the effects of candidate gender and age are lower than the effects of the remainder of the attributes.

Figure H: Effects of undemocratic behavior on candidate support across partisanship, policy distance, and preference intensity. Remaining candidate attributes (age, gender, competence reputations, and background) are included.



Note: Same model and sample as in original Figures 1-3 except that the remaining candidate attributes are included.

Figure I: Standard AMCE-plot showing effects of the included attributes on candidate support.



Note: OLS-regression with the support-variable regressed on the shown attributes. Policy distance, competence, and partisanship are run as linear effects. The coefficients shown for these variables, therefore, signal the effect of a one scale point change on the scales from 1-5. The employed sample is the exact same as in the original specifications.

Appendix J: Do Extreme Policy Preferences Matter for Sanctioning of Undemocratic Behavior?

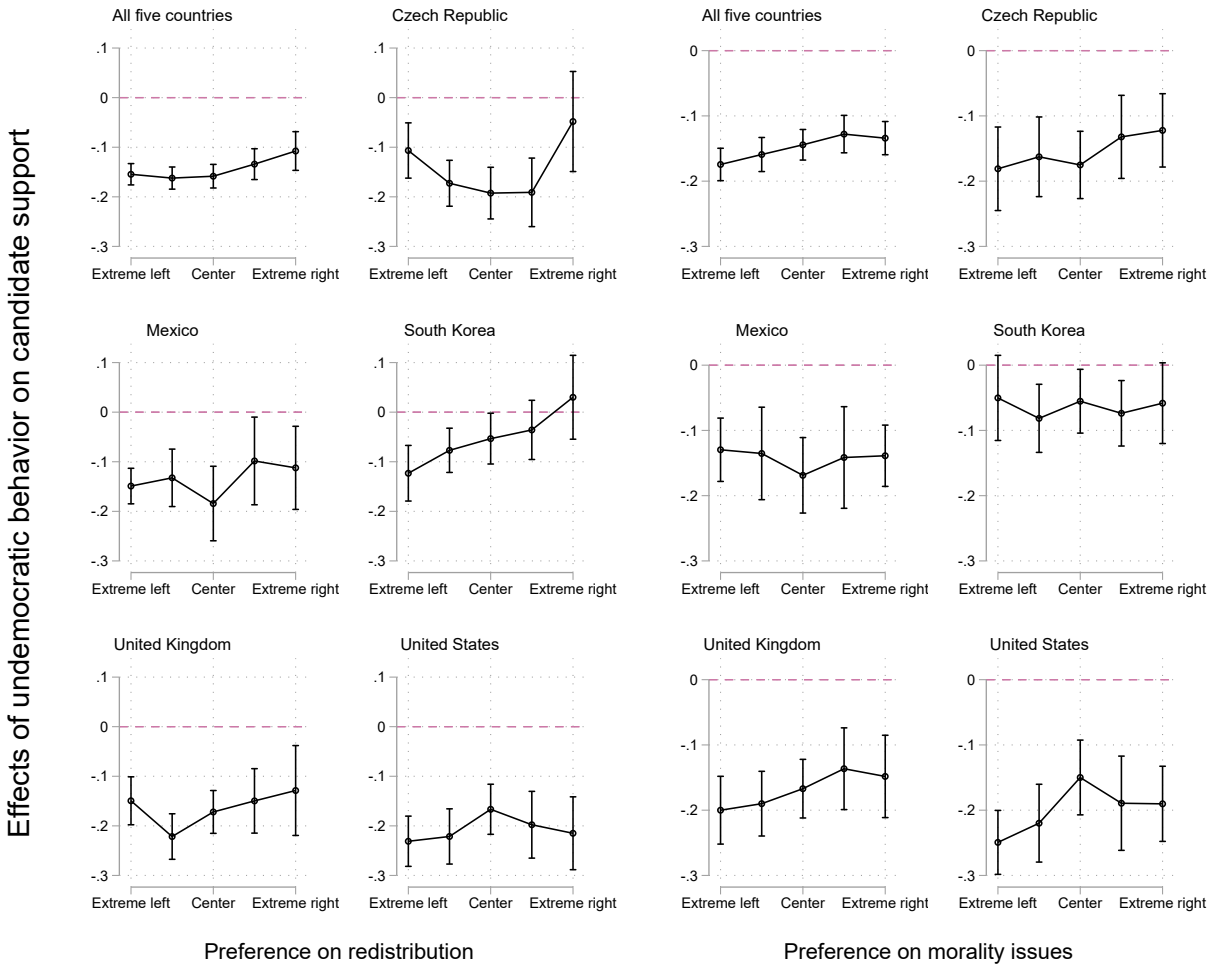
In the pre-registration, I hypothesized that extreme policy preferences as well as policy distance between voter and candidate should suppress the impact of undemocratic behavior. Based on comments and suggestions on presentations of this article, I decided to omit this part from the analysis. The reasoning is that it seems more interesting to look at how citizens sanction political opponents and allies as compared to just looking at whether extreme preferences in themselves matter.

Nevertheless, Figure J shows this omitted test by illustrating the effects of undemocratic behavior across preference extremity on redistribution and morality issues. Preference extremity is here measured by the five-point agree/disagree scale derived from answers to the statements shown in Table A4. On both issue dimensions, we see a slight tendency that sanctions for undemocratic behavior decrease from left to right: That is, right-wingers tend to sanction violations of democratic principles less than left-wingers. Moreover, we see no tendency that preference extremity in itself — that is, irrespective of direction — should matter for sanctioning of undemocratic behavior.

Appendix K: Assessing the Sensitivity of Results With New Experimental Data

In this appendix, I test the sensitivity of two findings using new experimental data. First, I assess the sensitivity of the main finding that partisan loyalty and policy interests do not diminish the effect of undemocratic behavior. Second, I assess the sensitivity of the secondary finding — also found elsewhere (Svolik 2020; Graham and Svolik 2020; Carey et al. 2020) — that voters are willing to compromise democratic compliance for partisan ends inferred from the observation that they prefer politically favorable undemocratic candidates over politically unfavorable democratic candidates. In short, I show that the main finding is robust across studies, while the secondary finding is very sensitive to the relative treatment effects of undemocratic behavior and political interests. Specifically,

Figure J: Effects of undemocratic behavior on candidate support across preference extremity.



Note: OLS-regression including an interaction term between undemocratic behavior and preference extremity with standard errors clustered on the respondent-level. Same sample as in the original specifications.

when effects of undemocratic behavior are larger, citizens seem far less willing to trade off democratic compliance for partisan ends.

The new data consist of two rounds of experimental studies which I conducted on Mechanical Turk in joint work with Associate Professor Lasse Laustsen (Aarhus University) in the US in March and May 2021. The research design is similar to the one which produced the main findings with minor deviations: That is, each respondent encountered ten candidate choice scenarios including candidate profiles varying in party, policy positions, and undemocratic behaviors (N = 450-500 respondents and 8,500-9,500 candidate observations in each round).

The minor deviations between studies are the following. First, the dependent variable was measured as a seven-point scale. Second, the measure of shared partisanship between respondent and candidate is based on standard ANES-procedures where respondents are asked which, if any, of the two parties they identify with categorically and then probed to reveal whether they are weak partisans, strong partisans, leaners, or true independents. This measure yields a seven-point scale of shared partisanship and is the exact same as the one used in Graham and Svolik (2020). Third, competence reputations were not included, which means that candidate profiles were slightly less complex than in the main study. Finally, three of the undemocratic behaviors are different in the second MTurk round. These new undemocratic behaviors are "Said elected officials should not be bound by court decisions they regard as politicized", "Said that election officials cannot be trusted", and "Said that use of force against certain groups can be necessary if they retain information about crimes". "Said it is acceptable to harass journalists who do not reveal sources" is included in all three studies.

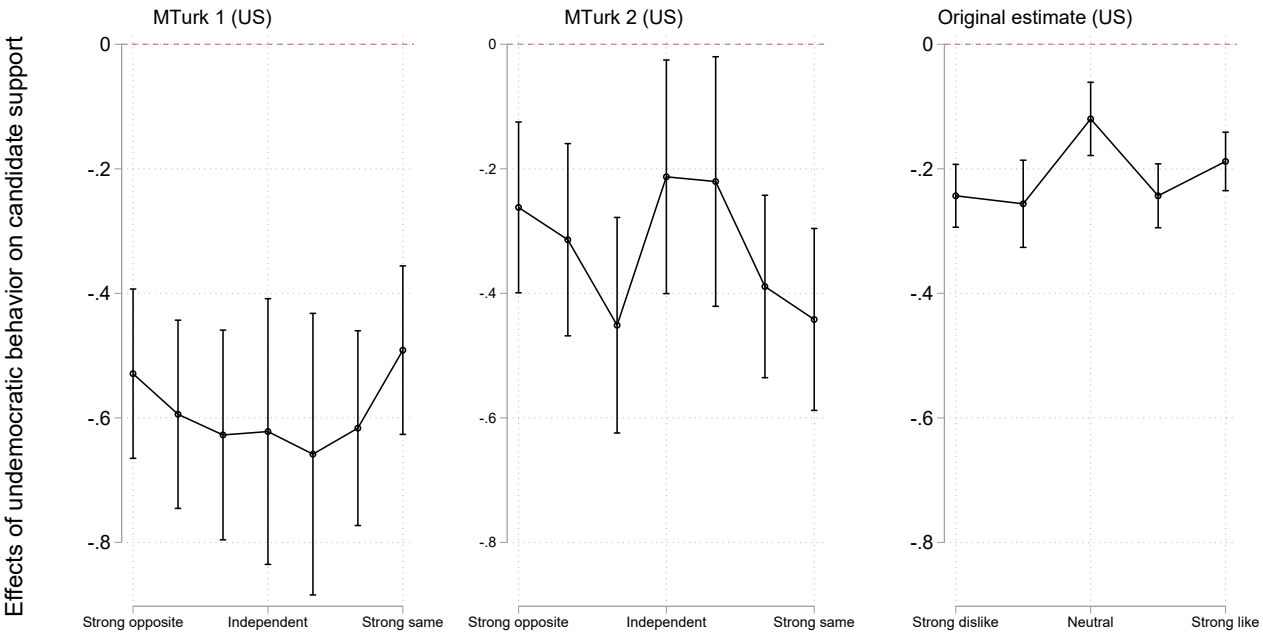
In the following figures, I have rescaled the dependent variables to 1-5 to ensure comparability. For the same purpose, I only show the US-estimates from the original study as the MTurk studies were conducted in this country. The figure pairs K1A-K1B, K2A-K2B, and K3A-K3B show the effects of undemocratic behavior (A-figures) and the levels of support for undemocratic and democratic candidates (B-figures) across partisanship (K1s), policy distance (K2s), and preference intensity measured as the extent

to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate (K3s).

Starting with K1A, we see that the finding that partisanship does not diminish the impact of undemocratic behavior holds across all three studies. Moreover, we see that the effects of undemocratic behavior differ: The effects are around twice as large in the first MTurk study as compared to in the original studies while the second MTurk study falls in between. Proceeding to Figure K1B, we see that the finding that voters are willing to compromise democratic compliance for partisan ends is very sensitive to which study we focus on: While voters prefer in-partisans regardless of democratic compliance in the original study, this is not the case in the two MTurk studies. In the first MTurk study, strongly partisan voters are as willing to support a democratic candidate from the other party as they are willing to support an undemocratic candidate from their own party. In all other comparisons, the democratic candidate wins irrespective of partisanship. In the second MTurk study, voters are slightly less willing to compromise their partisan loyalties but still far more willing than in the original study.

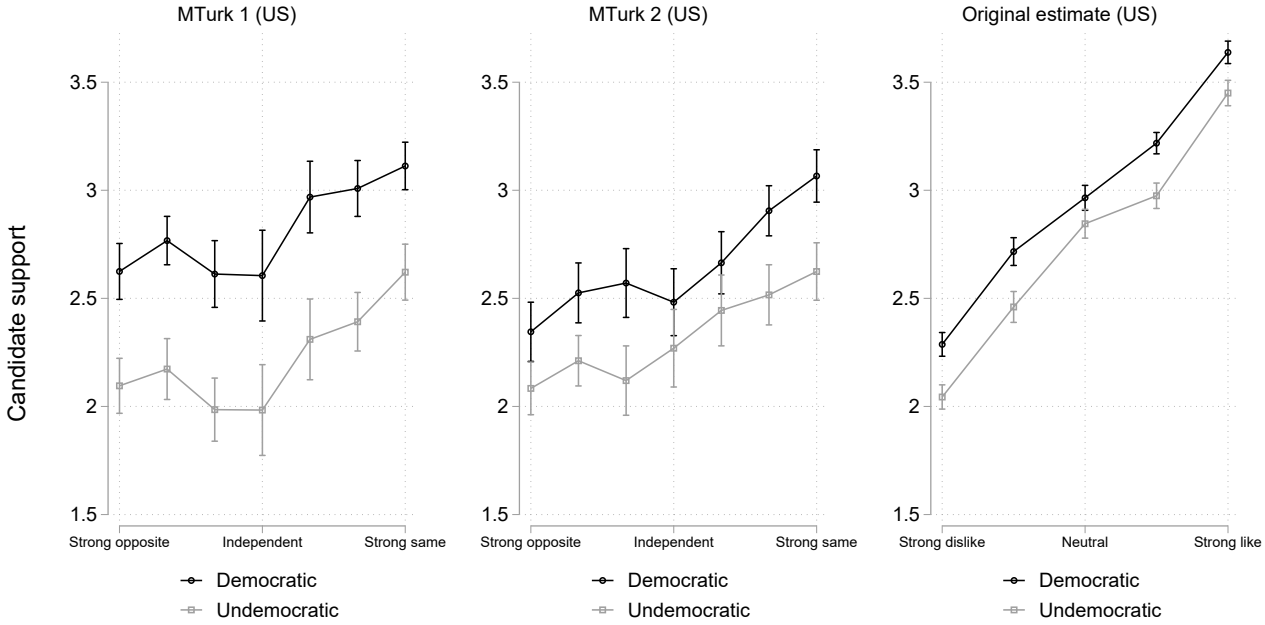
Figures K2A and K3A underline the finding that voters punish undemocratic candidates with whom they agree on policy the most. The moderating effects of policy agreement on the impact of undemocratic behavior are in fact even stronger in the two MTurk studies as compared to in the original study. Figures K2B and K3B first and foremost show that the impacts of policy agreement — just like that of undemocratic behavior — are stronger in the two MTurk studies as compared to in the original study. We can see this as the regression lines are steeper from left to right in the two MTurk-studies. This also means that — as both the impacts of undemocratic behavior and policy agreement are larger — voters' willingness to trade-off democratic compliance for policy interests is fairly similar across all three studies though slightly less pronounced in the first MTurk study. Yet again, this only underlines that the finding that voters prefer politically favorable undemocratic candidates over politically unfavorable democratic candidates is very sensitive to the relative strength of experimental treatments.

Figure K1A: Effects of undemocratic behavior on candidate support across shared partisanship between voter and candidate.



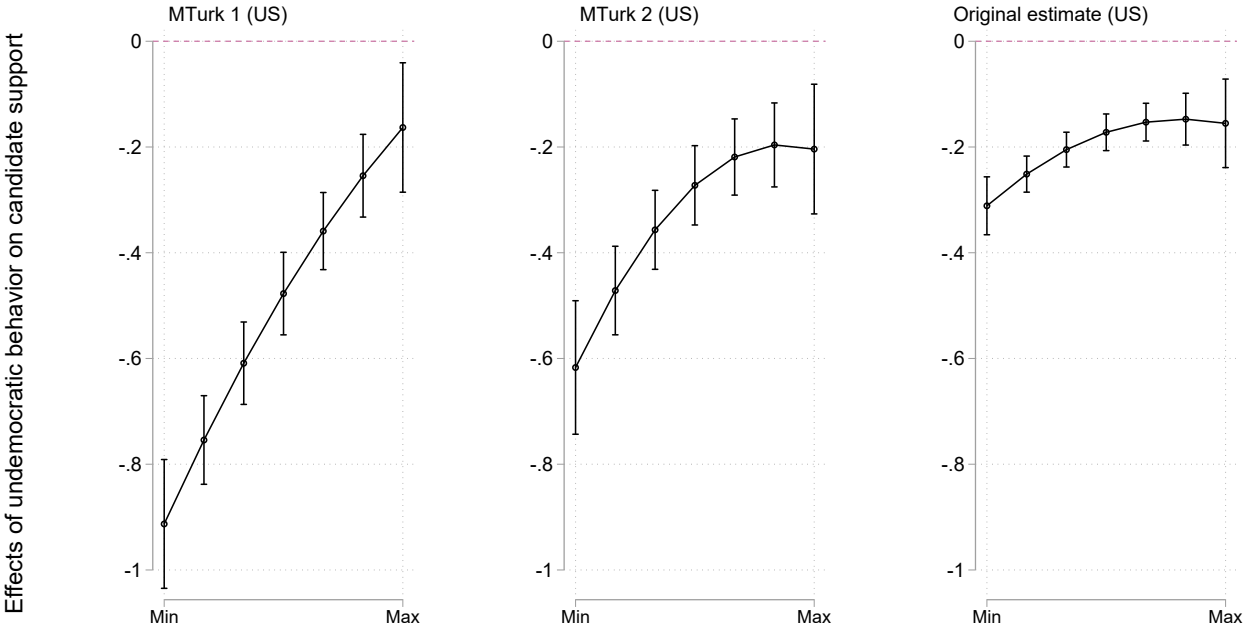
Note: Same model specifications as in Figure 1.

Figure K1B: Levels of support for undemocratic and democratic candidates across shared partisanship between voter and candidate.



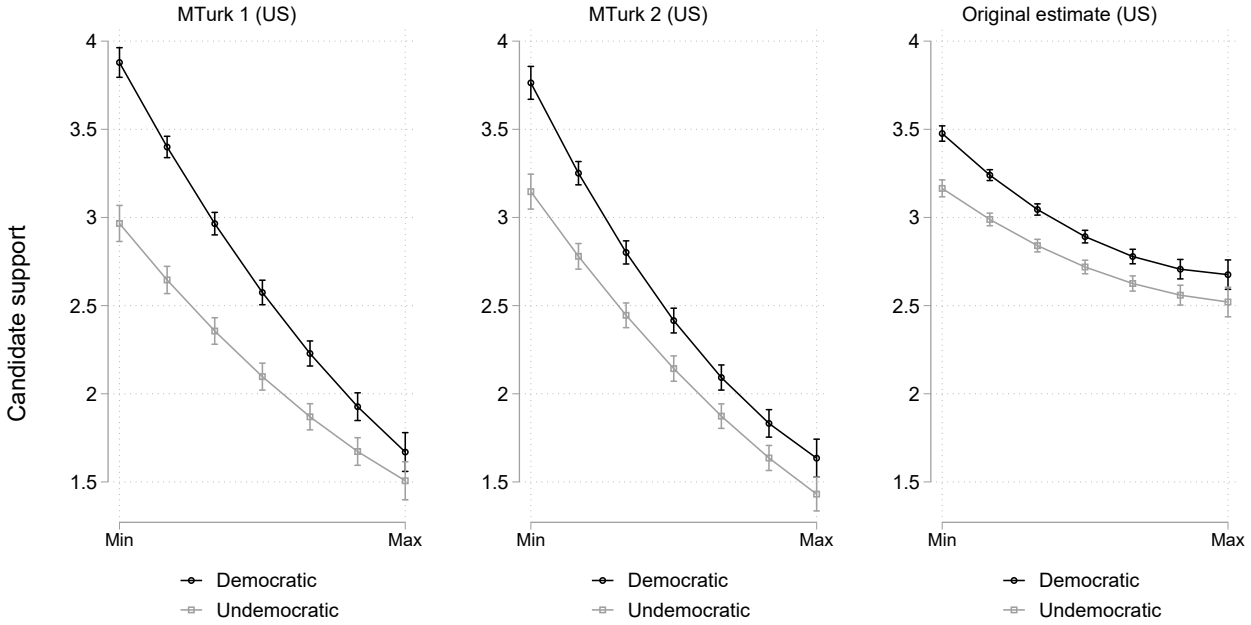
Note: Same model specifications as in Figure 1.

Figure K2A: Effects of undemocratic behavior on candidate support across policy distance between voter and candidate.



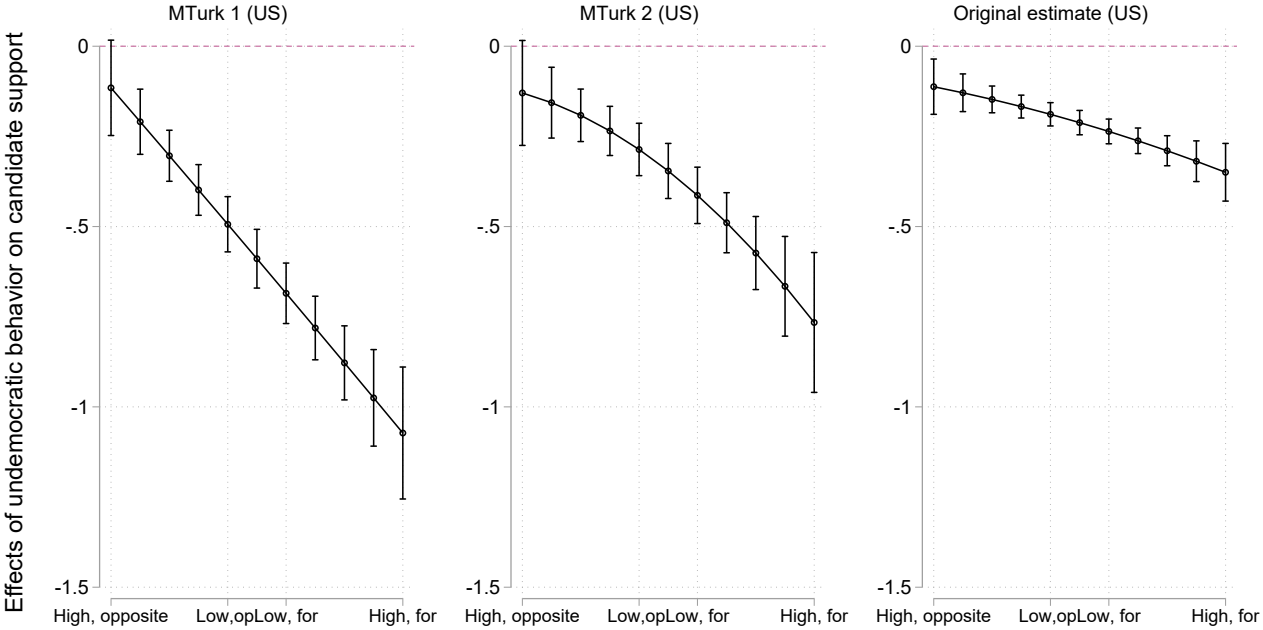
Note: Same model specifications as in Figure 2.

Figure K2B: Levels of support for undemocratic and democratic candidates across policy distance between voter and candidate.



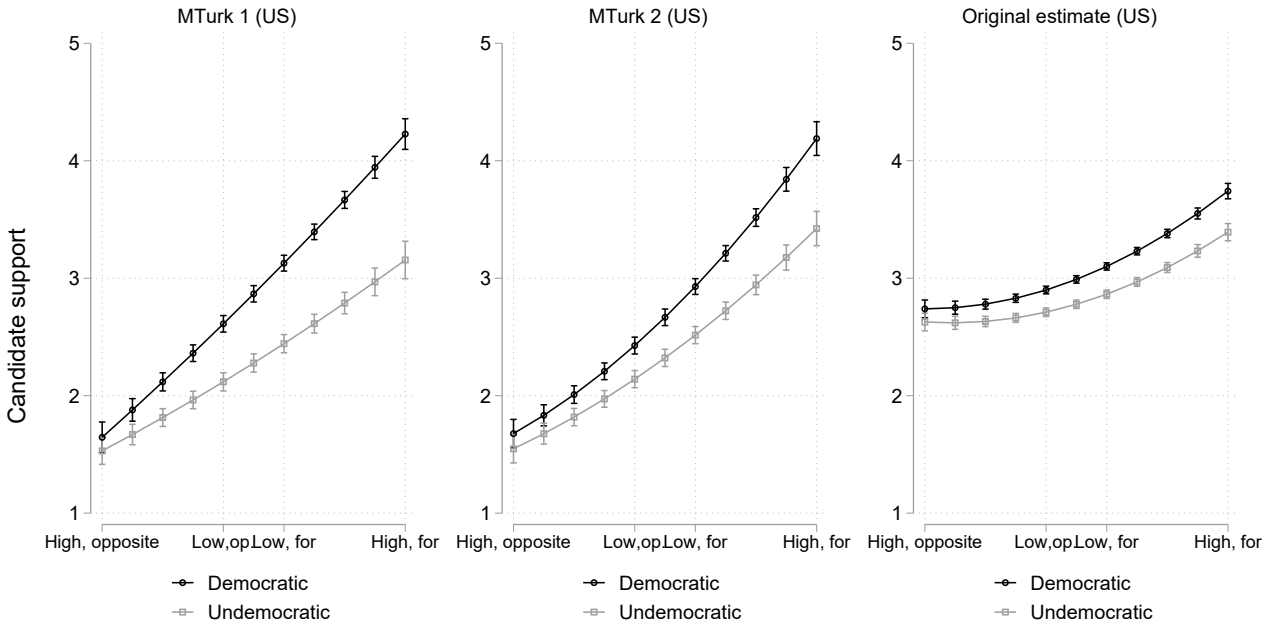
Note: Same model specifications as in Figure 2.

Figure K3A: Effects of undemocratic behavior on candidate support across preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate.



Note: Same model specifications as in Figure 3.

Figure K3B: Levels of support for undemocratic and democratic candidates across preference intensity measured as the extent to which the voter has intense or non-intense preferences in the same or opposite direction of the candidate.



Note: Same model specifications as in Figure 3.

In sum, the finding that citizens punish undemocratic behavior irrespective of partisan allegiance is very robust across studies, while the finding that citizens are willing to trade off democratic compliance for partisan ends depends heavily on the relative treatment effects of undemocratic behavior and political interests. A likely explanation for why the MTurk studies produced larger effects of undemocratic behaviors is that MTurk-respondents are more politically aware and thus perceive undemocratic behaviors as more severe than Lucid-respondents do (Coppock and McClellan 2019: 6).

Appendix L: Description of Deviations from Pre-registration

In this appendix, I list each deviation from how the study is described in the pre-registration accessible at https://osf.io/qjm42. As also mentioned in the article, this entire study is only a smaller part of a larger project outlined by the pre-registration (specifically testing H3a, H3b, and H3c). On the basis of suggestions and comments on the project collected after conducting the pre-registration, I decided to split up the reporting of the different hypotheses to keep all parts fairly concise.

1. **Reporting/wording of H3b and H3c.** In the pre-registration, H3b is about mere preference extremity as well as policy distance between citizens and candidates. I have moved the part about preference extremity to the appendix. See Appendix J for a justification of this choice and for the actual test of whether preference extremity in itself matters for sanctioning of undemocratic behaviors. Moreover, I have specified H3c slightly in the article as compared to in the pre-registration ("Intense policy preferences — to the same or opposite side of the candidate — diminish the negative effect of undemocratic behavior" versus "Intense policy preferences diminish the negative effect of undemocratic behavior"). I made this adjustment to clarify that I take the candidates' policy positions into account when testing this hypothesis. It is evident from the pre-registered empirical tests of the hypothesis that I would do this, so the lack of specification in the hypothesis was simply an error in the pre-registration.

2. **The sample.** The samples of each country employed in the article are larger than the approximation in the pre-registration. The pre-registration states that the samples would be approximately 1,500-1,800 respondents in each country whereas the actual samples are between 2,350 and 2,999 respondents. The reason for this discrepancy is simply that it became evident during the data collection that it was possible to collect more data within the relevant limitations.

3. **Execution of analysis.** I have implemented a few minor deviations from the analysis plan evident from the text and do-file in the pre-registration.

   In the analysis do-files attached to the pre-registration, I planned to estimate the interaction between undemocratic behavior and the preference intensity measure separately for intensity on redistribution and morality issues, respectively. In the article, I have summed the two intensity measures into one to keep the test more concise and limit the number of figures. I show in Appendix C that the findings are not sensitive to this choice.

   Moreover, while the pre-registration states that I would use partisanship in its squared form when estimating the interaction with undemocratic behavior, I estimate the interaction non-parametrically (that is, I employ partisanship as a factor variable) in the article. I simply do this because the interaction did not turn out to have any clear functional form. As I suspected that this might be the case when writing up the pre-registration, I did in fact specify this non-parametric estimation in the do-file as an alternative option in case the functional form of the interaction did not turn out as predicted. Nonetheless, I show in Appendix F that using partisanship in its squared form does not change the findings.

References

Carey, J., K. Clayton, G. Helmke, B. Nyhan, M. Sanders, and S. Stokes (2020). Who will defend democracy? evaluating tradeoffs in candidate support among partisan donors and voters. *Journal of Elections, Public Opinion and Parties 0*(0), 1–16.

Coppock, A. and O. A. McClellan (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research and Politics 6*(1), 1–14.

Graham, M. H. and M. W. Svolik (2020). Democracy in America? partisanship, polarization, and the robustness of support for democracy in the United States. *American Political Science Review 114*(2), 392–409.

Svolik, M. W. (2020). When polarization trumps civic virtue: Partisan conflict and the subversion of democracy by incumbents. *Quarterly Journal of Political Science 15*(1), 3–31.