Fixed Effects and Post-Treatment Bias in Legacy Studies*

Jonathan Homola
University of California, Los Angeles
homola@ucla.edu

Miguel M. Pereira
London School of Economics and Political Science
m.m.pereira@lse.ac.uk

Margit Tavits
Washington University in St. Louis
tavits@wustl.edu

Abstract

Pepinsky, Goodman, and Ziller (PGZ 2023) reassess a recent study on the long-term consequences of concentration camps in Germany (Homola, Pereira, and Tavits 2020). The authors conclude that accounting for contemporary (i.e., post-treatment) state heterogeneity in the models provides unbiased estimates of the effects of camps on current-day outgroup intolerance. In this note we show that PGZ's empirical strategy rests on (a) a mischaracterization of what regional fixed effects capture and (b) two unrealistic assumptions that can be avoided with pre-treatment state fixed effects. We further demonstrate that results from the original article remain substantively the same when we incorporate regional fixed effects correctly. Finally, simulations reveal that camp proximity consistently outperforms spatially correlated noise in this specific study. The note contributes to the growing literature on legacy studies by advancing the discussion about the correct modeling choices in this challenging field.

*We would like to thank Avidit Acharya, Volha Charnysh, Vicky Fouka, Anselm Hager, Soeren Henn, Florian Hollenbach, Connor Huff, Christopher Lucas, Noam Lupu, Jacob Montgomery, Leonid Peisakhin, Arturas Rozenas, Anton Strezhnev, Michelle Torres, Rick Wilson, and Yuri Zhukov for comments and helpful discussions about earlier versions of the paper.

A prominent set of studies in political science shows that long-deceased coercive institutions often continue to influence contemporary political attitudes and behavior (e.g., Acharya et al. 2016a; Lupu and Peisakhin 2017; see Simpser et al. 2018 and Charnysh et al. 2023 for reviews). Reliably establishing legacy effects is challenging. It requires making robust causal inferences over a very long time span during which the treatment could have affected not only the outcome but also other variables relevant for the analysis. To deal with this challenge, most legacy studies explicitly address post-treatment bias by employing appropriate methods such as the sequential g-estimator.

Building upon the analytical strategies established in the legacies literature, Homola et al. (2020 – HPT) explore the long-term political consequences of the Third Reich. The results show that current-day political intolerance, xenophobia, and voting for radical right-wing parties are associated with proximity to former Nazi concentration camps in Germany. This conclusion relies on a series of analyses using election results and data from two different surveys to measure contemporary attitudes and behavior.

Pepinsky et al. (2023 – PGZ) re-examine HPT and argue that state-level differences confound the relationship between distance to camps and out-group intolerance. To overcome this issue, PGZ add contemporary state fixed effects to HPT's models and find that proximity to concentration camps is no longer a reliable predictor of intolerance. The authors posit that contemporary states, although mainly formed after the Third Reich, do not introduce post-treatment bias if the following assumptions hold: (a) contemporary cross-state heterogeneity is not in the causal path between camp proximity and contemporary attitudes, and (b) there are no unobserved variables that jointly explain contemporary state differences and contemporary outgroup intolerance or camp proximity.

We agree that it is important to think carefully about spatial heterogeneity in the historical legacies literature. HPT's original analysis accounts for regional heterogeneity by including controls such as the local-level share of unemployment and foreigners, urban status, or a dummy for East vs. West Germany. While HPT considered these solutions sufficient and did not find a theoretical motivation to include state fixed effects, other scholars operating in good faith might find doing so important for theoretical or empirical reasons.

We also agree with the authors that adding variables observed post-treatment to a model does not *always* bias the estimates. However, PGZ's conclusion that current-day state fixed effects do not risk inducing post-treatment bias rests on a dual and inconsistent interpretation of what regional fixed effects capture.

The authors emphasize the importance of accounting for regional heterogeneity by noting that "Länder fixed effects adjust for any factor (observable or not) that varies across German Länder" (p. 2). Any political, economic, or social dynamic that varies across states is captured by geographical fixed effects. However, the meaning of this same construct shifts when PGZ describe the conditions for contemporary state fixed effects to induce post-treatment bias: "unless distance to concentration camps (T) causally affects postwar Länder boundaries (F), controlling for Länder fixed effects cannot create posttreatment bias" (p. 3). This statement is incorrect. Following the definition that PGZ used earlier, contemporary state fixed effects induce post-treatment bias if "any factor (observable or not) that varies across German Länder" is a direct or indirect descendant of proximity to concentration camps. In other words, we need to assume that Nazi concentration camps had no effects whatsoever that vary systematically across states. Even with a perfectly random geographic distribution of the camps across Germany, this assumption only holds if camps had no effects at all on the economic and social structure around them – something that is not consistent with existing evidence (Charnysh and Finkel 2017; Hoerner et al. 2019). In brief, the very reason why PGZ emphasize that state fixed effects are important is also the reason why we should be concerned about post-treatment bias.[1]

In the remainder of this note, we first discuss the plausibility of the assumptions invoked by PGZ to support their empirical strategy. Next, we describe how it is possible to account for

[1]Additionally, PGZ's argument rests on a second implausible assumption: that any predictor of outgroup intolerance, such as economic conditions or cultural resentment, *does not* vary systematically by state, and that there is no unexplained variation driving camp locations and contemporary state heterogeneity.

state-level heterogeneity in HPT's analyses without risking post-treatment bias or M-bias and without requiring PGZ's strong assumptions. By replacing PGZ's contemporary state fixed effects with Weimar-era state fixed effects – the state boundaries in place at the time when the first camps were built – the results in HPT remain substantively unchanged. Finally, in the Supplementary Materials (SM) we assess the robustness of the findings to spatial correlation by replacing the geographic variable in HPT with spatially correlated noise, an appropriate alternative to assess spatial autocorrelation (Kelly 2019). The simulations suggest that camp proximity consistently outperforms spatial noise as an explanatory variable.

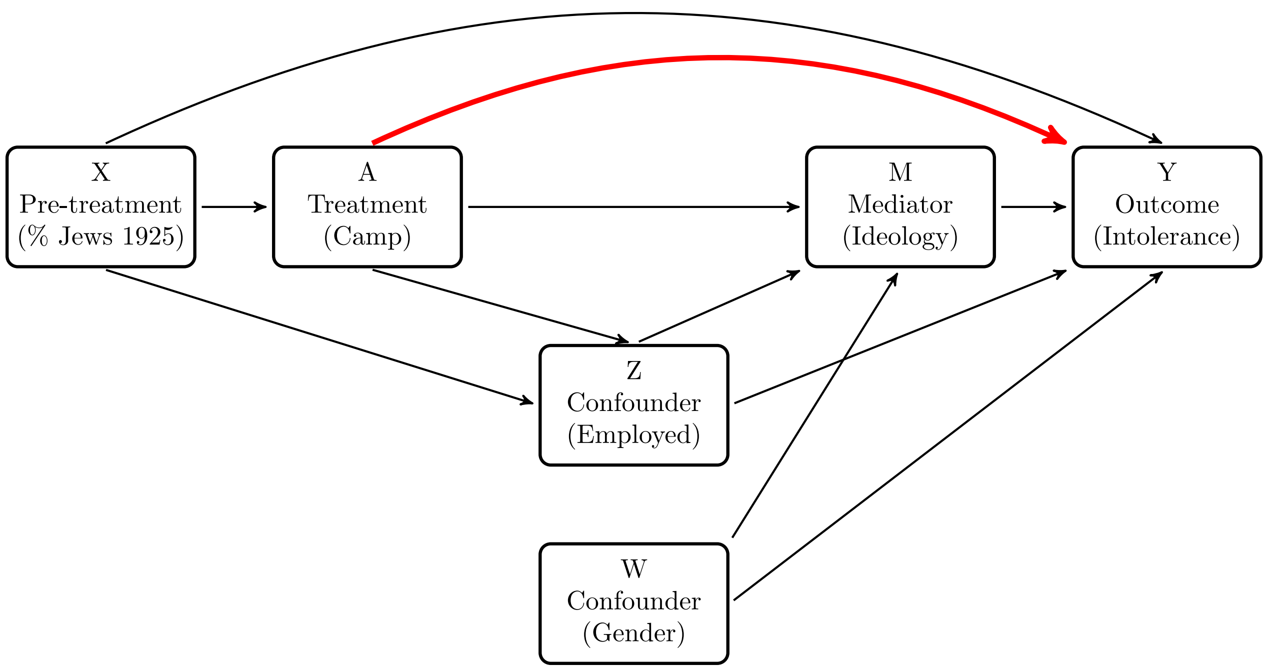
Taken together, the note contributes to the growing literature on legacy studies by offering practical solutions for correctly dealing with regional heterogeneity in this challenging field. It also highlights the importance of thinking carefully about what statistical tools and concepts (e.g., fixed effects) represent and account for, as well as their underlying assumptions.

Post-treatment bias and how to avoid it

Post-treatment bias occurs when an analysis conditions on a variable that is directly or indirectly affected by the treatment and also shares a common cause with the outcome of interest. Post-treatment bias is especially problematic because without very strong assumptions, it is impossible to know how it affects our estimates of interest. Neither the direction nor the magnitude of the bias are possible to anticipate (Elwert and Winship 2014; Montgomery et al. 2018).

To overcome these difficulties in settings where we have strong theoretical reasons to include variables that are measured post-treatment, complex modeling strategies and additional assumptions are necessary. The sequential g-estimator is one such approach (Acharya et al. 2016b). The method starts by estimating a model with pre-treatment and post-treatment covariates (1st stage). Next, it recalculates the outcome variable by removing from it the effects of the mediating variables of interest. Finally, it estimates the effect of the treatment on this "demediated" outcome (2nd stage). The model allows us to incorporate post-treatment confounders and mediators without incurring in post-treatment bias.

Figure 1: Directed acyclic graph illustrating the g-estimator



Note: The bold red line represents the controlled direct effect. Pre-treatment (i.e., Weimar-era) state-level fixed effects can be included as pre-treatment variables X. Post-treatment (i.e., contemporary) state-level fixed effects can be included as post-treatment confounders Z or W.

Figure 1 shows a causal relationship between different types of variables of interest, where the flow of causality runs from left to right. We have a set of pre-treatment variables X, the treatment variable A, a mediator M, and the outcome of interest Y. In addition, there are two types of confounders: Z, which is affected by the pre-treatment variables and the treatment, and W, which is not. Both Z and W confound the relationship between the mediator and the outcome.

In HPT, the authors use variables like a district's share of Jews in 1925 or unemployment rate in 1933 as pre-treatment variables X. The treatment A is the distance of a survey respondent (or area in the analyses with electoral data) to the closest former concentration camp, and the outcome of interest Y is captured by different indicators of out-group intolerance. We can think of the control variables (e.g., a respondent's employment status) as confounders Z that are likely to be affected by X and A. Other controls, such as age or gender, can be thought of as confounders W that are unlikely to be affected by X and A. Similarly, we can think of an individual's ideology as a mediator M.

To analyze this data structure, HPT employ the sequential g-estimator (Acharya et al. 2016a; 2016b; Robins et al. 1992). PGZ introduce contemporary state fixed effects into this setup. As recognized by the authors, decades passed between the construction of the camps and the creation of the German states we know today. Only six of the sixteen current states already existed when the first camp was built.[2] PGZ's decision to control for post-treatment state-level heterogeneity rests on two assumptions that we critically evaluate below.

PGZ's assumption 1: contemporary state heterogeneity is not explained by camp locations

The first assumption for PGZ's analyses to hold is that any contemporary cross-state differences (in attitudes, socioeconomic conditions, economic development, etc.) are not explained directly or indirectly by the location of camps. We rely on three pieces of qualitative evidence to demonstrate why this assumption is unlikely to hold.

First, after World War II, southwestern Germany initially consisted of three states: Baden, Württemberg-Hohenzollern, and Württemberg-Baden. The Württemberg-Baden government wanted to unify all three into a single state, but Baden was against it. The new Basic Law from 1949 contained a specific article, which clarified that if the states could not come to an agreement, a referendum would be held. This referendum took place on December 9, 1951 and ultimately resulted in a merger of the three states into the new state of Baden-Württemberg.[3] The political discussion in the run-up to the referendum focused on economic and administrative issues but also on out-group resentment, including anti-

[2] In Appendix SM 2 we describe in detail how current-day states differ from the Weimar-era states.

[3] `https://www.lpb-bw.de/entstehung-baden-wuerttembergs`

Baden attitudes and religious factions (Weber and Häuser 2008). In other words, in this specific instance citizens themselves determined the shape of their states. If the concentration camps affected people's beliefs during the Third Reich, as HPT argue (see also Charnysh and Finkel 2017; Hoerner et al. 2019), and these same people decided the shape of states created post-war, then contemporary state differences are directly in the causal path between camp locations and current-day attitudes (as M or Z in Figure 1, above).[4]

Second, and most importantly, *contemporary states can induce post-treatment bias even without the direct or indirect influence of camps on state borders.* Another example highlighted by PGZ helps illustrate this. The authors mention one specific difference across states that their fixed effects capture: the existence of "variation in school curricula" (p. 2). While all state curricula include the discussion of the Nazi regime, there is systematic variation in whether or not students visit a concentration camp. This variation is driven in part by proximity to a camp. Schools are more likely to organize a camp visit if there is a camp close by. Some states even subsidize camp visits if they happen within the same state (Rathenow and Weber 1995; see Fouka and Voth 2023 for similar evidence in Greece). Therefore, policies determined by today's states are shaped by proximity to camps and affect the likelihood that students will visit a camp. Contemporary state fixed effects, if treated as a pre-treatment confounder, pick up these state-level differences and induce post-treatment bias.

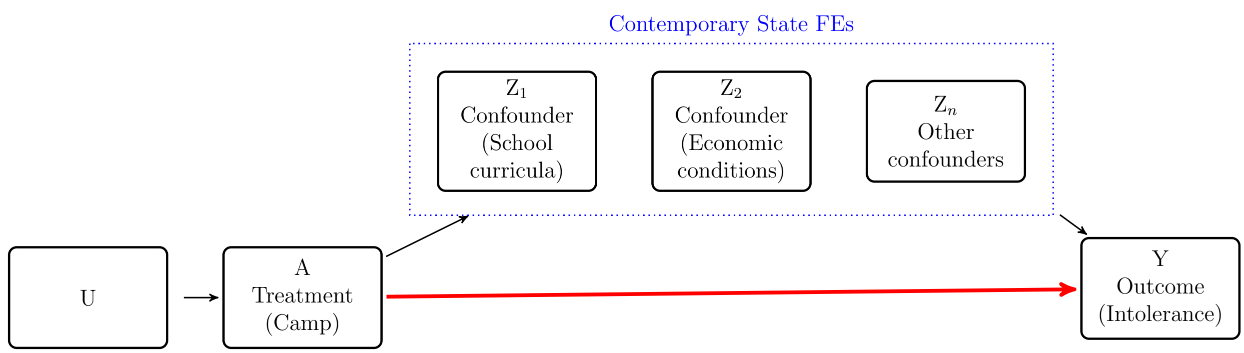
Finally, on a more abstract level, we know that not every state has the same number of camps. For example, Thuringia has two camps in HPT's analysis although it is among the smallest states in Germany. On the other hand, North Rhine-Westphalia is one of the largest states but does not have any camps. If we assume that the camps had any effect at all on their surrounding areas, and we know that some small states have multiple camps whereas some large states have no camps, then these camp effects are necessarily leading to state-level

[4] None of the main concentration camps was located inside Württemberg-Baden. Our argument is meant to illustrate ways through which individual citizens directly helped reshape administrative borders after the end of the Third Reich.

differences that are (a) clearly post-treatment, and (b) would be picked up by fixed effects. The results of Charnysh and Finkel (2017) demonstrate exactly this: the area surrounding the Treblinka camp in Poland experienced a real estate boom following the closure of the camp, and local communities in the area are subsequently more supportive of an anti-Semitic party. In other words, the camp had attitudinal effects that were concentrated in the region surrounding the camp and would be picked up by contemporary state fixed effects.

Figure 2 expands a section of Figure 1 to illustrate our points. This DAG corresponds to Figure 1b in PGZ. Contemporary fixed effects are likely to capture a collection of post-treatment confounders which induces post-treatment bias.

Figure 2: Violation of PGZ's assumption 1



Note: Adaptation of PGZ's Figure 1b showing a violation of the assumption that any contemporary cross-state differences are not explained directly or indirectly by the location of camps. The bold red line represents the causal effect of interest.

PGZ's assumption 2: no collider bias

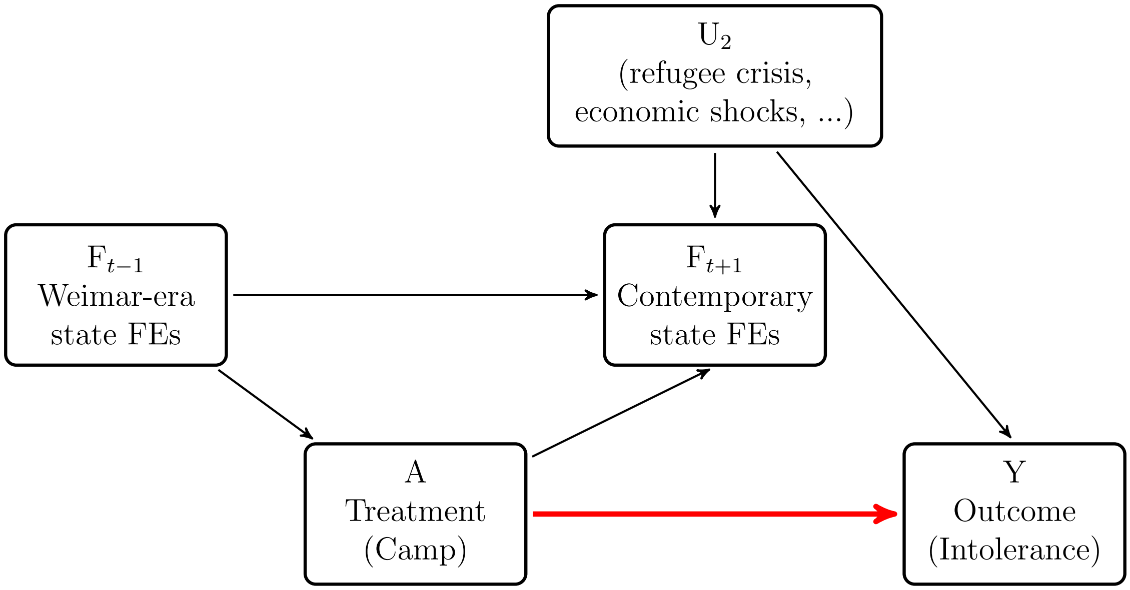
PGZ's empirical strategy relies on a second assumption: that contemporary state differences are not explained simultaneously by (a) a variable that also predicts contemporary outgroup attitudes (the outcome) and (b) a variable that predicts camp proximity (the treatment). Violating this assumption leads to a form of collider bias (M-bias) and produces spurious causal inferences.

PGZ defend this assumption by stating, "there is no interpretation of German administrative history that matches any of the hypothetical causal structures [described by the authors]" (p. 6). Once again, PGZ reduce state fixed effects to a matter of administrative borders. However, once we acknowledge what regional fixed effects capture, this assumption becomes considerably less plausible.

First, virtually all of the canonical predictors of outgroup intolerance in Germany are not homogeneously distributed across the territory but instead geographically clustered in certain areas and can therefore also predict cross-state differences. These include economic insecurity (Funke et al. 2016), perceived cultural threat (Norris and Inglehart 2018), distribution of globalization "winners" and "losers" (Kriesi et al. 2006), and perceived security threat (Ward 2019). Second, there is, in fact, a plausible predictor of contemporary state differences *and* camp proximity: the Weimar-era states. As PGZ describe, the states in place prior to the creation of the camps overlap with contemporary states at least partly. Additionally, the geographical clustering of the camps described above means that Weimar-era states predict camp proximity. Figure 3 puts this in formal terms using the DAG presented by PGZ in Figure 2b. Weimar-era states (F_{t-1}) are a predictor of the treatment (A) and PGZ's fixed effects (F_{t+1}); in turn, conventional predictors of outgroup intolerance (U_2) influence contemporary state heterogeneity and the outcome. This violates the implausible and untestable assumption that F_{t+1} is not a collider.

These examples reveal the challenges of making assumptions about causal structures with variables that capture many sources of variation, such as regional fixed effects. By introducing in the causal model an amorphous cluster of variables that cannot be isolated, PGZ are forced to rely on untestable and implausible assumptions. Next, we describe two solutions to avoid the assumptions invoked by PGZ: (a) correctly specifying contemporary states as post-treatment variables, or (b) using pre-treatment state fixed effects.

Figure 3: Violation of PGZ's assumption 2



Note: Adaptation of PGZ's Figure 2b showing a violation of their assumption that there are no predictors of contemporary state differences (F_{t+1}) and camp proximity (A). The bold red line represents the causal effect of interest.

Two solutions to account for regional heterogeneity

Consider again the causal relationship in Figure 1. PGZ treat contemporary state fixed effects as pre-treatment variables X. Above, we described different ways in which this modelling choice induces post-treatment bias. However, g-estimation provides a way to overcome these flaws. Instead of treating contemporary states as pre-treatment variables, we consider them to be post-treatment confounders W or Z. This assumes that contemporary states are confounding the relationship between the mediator (e.g., ideology) and the outcome (e.g., out-group intolerance), which makes intuitive sense and is in line with the potential confounding effects of state heterogeneity that PGZ discuss ("unsynchronized policy environments," "substantial variations in school curricula"). HPT adopt this same procedure to account for systematic differences between East and West Germany.

In terms of the estimation, this means that contemporary state fixed effects should appear in the first stage of the sequential g-estimator, *but not in the second stage.* Recall that the goal of the first stage is to accurately estimate the effect of the mediator on the outcome to successfully "de-mediate" the outcome before the second stage. The confounders W or Z (i.e., the contemporary state fixed effects) are only relevant for this part of the estimation and should not be included in the second stage.

An alternative solution to avoid post-treatment bias involves replacing contemporary state fixed effects with pre-treatment state fixed effects. We use Germany's administrative map from 1932 as seen in Figure SM2.1 to identify the corresponding Weimar-era state for each present-day geographic location. We chose 1932 because it is the year before the first German camp was created (Dachau, March 1933). Theoretically, the use of the Weimar states means that we are now working with true pre-treatment variables X (cf. Figure 1). The Weimar states might affect camp locations (the treatment) through their policy environment or other unobserved factors that the other pre-treatment variables did not capture. They can also affect some of the contemporary confounders (Z) and the outcome variables (Y). Empirically, it means that the Weimar state fixed effects can now be included in both stages of the g-estimator without inducing post-treatment bias.[5] Crucially, because these are now pre-treatment variables, we do not have to make *any* assumptions about how they might be affected by the treatment.

Finally, we combine both approaches by including contemporary state fixed effects in the first stage and Weimar-era state fixed effects in both stages of the g-estimator. This third specification allows us to simultaneously account for historical regional differences that may explain camp location, and for any post-treatment confounder that varies systematically across contemporary states.

Together, HPT, PGZ, and the current note provide an extensive list of models with different data sources and model specifications. To help the reader follow this collective

[5]PGZ also include an analysis with Weimar-era fixed effects, but they treat Prussian internal provinces as separate states. As a result of this further geographical slicing, the effects of camp proximity are no longer reliable. In SM 5 we discuss these analyses in greater detail and explain that the decision to split Prussia into provinces is arbitrary and atheoretical. Additionally, we show that when we include province-level fixed effects in the electoral analysis, HPT's main findings remain unchanged.

effort, Table SM1.1 summarizes the different main specifications modeling the effect of camp proximity on contemporary outcomes.

Results

We replicate the main analyses in HPT, i.e., the results in Table 2 (EVS) and Table 4 (electoral data) while including (a) contemporary state-level fixed effects in the first stage of the g-estimator, (b) Weimar-era state-level fixed effects in both stages of the g-estimator, and (c) both contemporary states *and* Weimar-era states.[6]

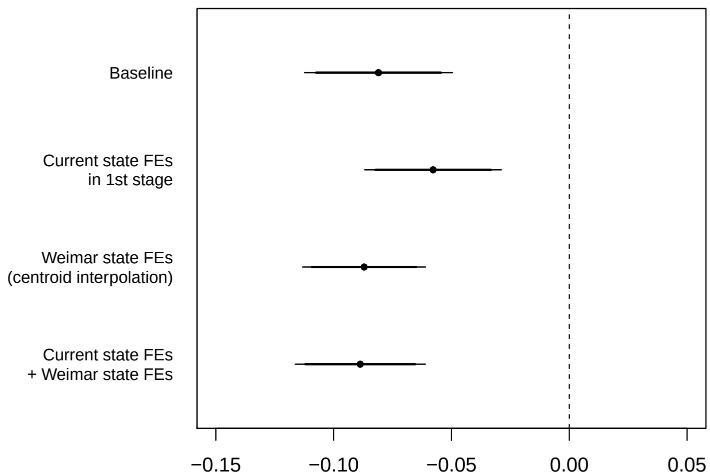
Figure 4 replicates the four columns of Table 4 in HPT. Each panel shows the controlled direct effect of camp proximity on support for radical right parties in 2017 for four different model specifications. The first coefficient (*Baseline*) is the effect reported in HPT. The second coefficient (*Current state FEs in 1st stage*) reports the results from the same model specification when we also include contemporary state fixed effects in the first stage of the g-estimator. The third coefficient (*Weimar state FEs*) corresponds to the models including Weimar-era state fixed effects in both estimation stages. Finally, the last coefficient (*Current state FEs + Weimar state FEs*) corresponds to models simultaneously accounting for contemporary regional differences in the first stage and Weimar-era fixed effects in both stages.

The results show that the main conclusions in HPT are robust to the inclusion of state fixed effects. Across the different specifications, we see that the effect of distance is always negative and statistically reliable at conventional levels except for the models within a 70km radius and with pre- and post-treatment regional fixed effects.[7] We also do not observe any dramatic change in the uncertainty of the estimates.

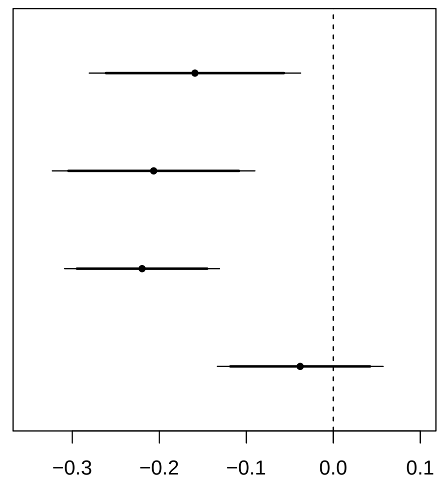
[6] The analyses do not include ALLBUS data because access to the dataset requires an in-person visit to the GESIS facilities in Cologne.

[7] We originally used the models focusing on a 70km radius around each camp because they lead to samples that are better matched in terms of potential confounders. However, this approach includes dropping over 60% of all data points and consequently also implies a loss of variation in the treatment. The models with pre- and post-treatment fixed effects include a total of 33 regional fixed effects. Given the already restricted sample, it is therefore not surprising that the effects are no longer statistically reliable at conventional levels.

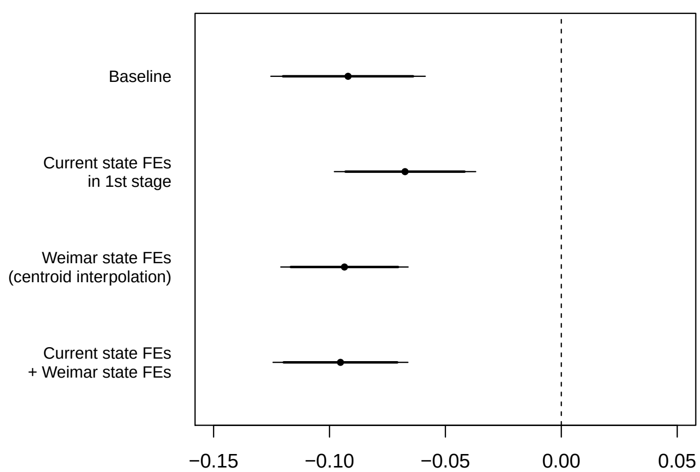
Figure 4: The controlled direct effect of camp proximity on support for radical right parties in 2017, accounting for state-level heterogeneity



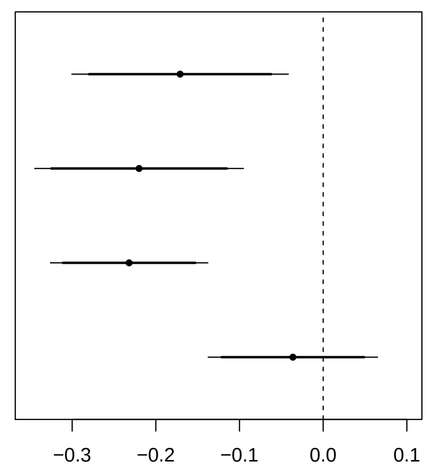
(a) AfD, full sample



(b) AfD, 70km radius



(c) AfD+NPD, full sample



(d) AfD+NPD, 70km radius

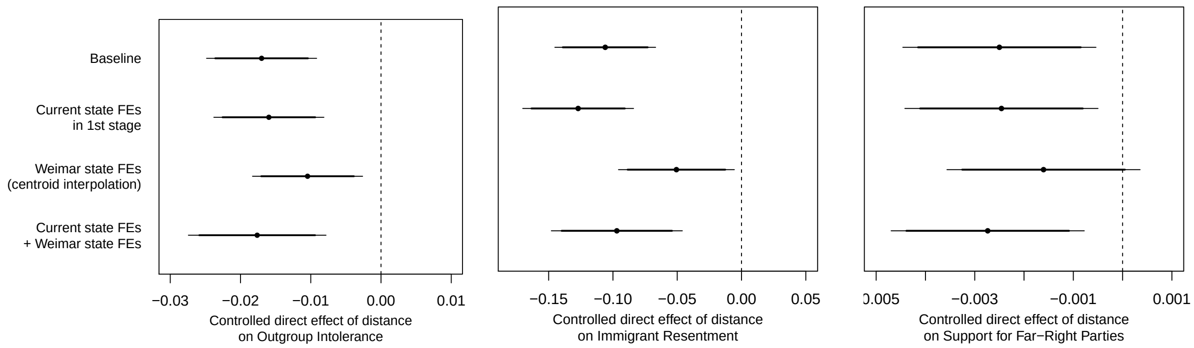
Note: Plots depict estimates and 95/90% confidence intervals from the sequential g-estimator for the controlled direct effects of distance to camps on support for radical right parties in 2017 (described in each panel label). Each estimate corresponds to a different model specification, described on the y-axis. The baseline specification corresponds to the results reported in Table 4 in HPT (N = 10,755 (a, c) and 3,949 (b, d)). Full model results for the remaining specifications in Tables SM3.1-SM3.3.

The three panels in Figure 5 repeat this exercise for HPT's EVS analysis. More specifically, each panel replicates the results in HPT, Table 2, Columns 2, 4, and 6, respectively, corresponding to a different outcome variable.[8] We report the controlled direct effect of

[8]Figure SM4.2, in the Appendix, replicates Columns 1, 3, and 5 from the same table, with OLS models.

camp proximity for the same four model specifications as in Figure 4.

Figure 5: The controlled direct effect of camp proximity on outgroup intolerance, immigrant resentment, and support for far-right parties (EVS), accounting for state-level heterogeneity



(a) Outgroup Intolerance **(b)** Immigrant Resentment **(c)** Far-Right Support

Note: Plots depict estimates and 95/90% confidence intervals from the sequential g-estimator for the controlled direct effects of distance to camps on contemporary attitudes (described in each panel label). Each estimate corresponds to a different model specification, described on the y-axis. The baseline specification corresponds to the results reported in Table 2 (Models 2, 4, and 6; N = 1,376) in HPT. Full model results for the remaining specifications in Tables SM4.1-SM4.3.

Again we find support for the original conclusions. Across the different outcome variables and fixed effects specifications, we see that the effect of distance is always negative and reliable. When we introduce contemporary state fixed effects as post-treatment variables, the main results are virtually unchanged. When Weimar-era fixed effects are introduced, the effect sizes decrease slightly in all models. However, for both attitudinal measures (panels a and b) the effects remain statistically significant. Only the estimate for self-reported far-right support (panel c, coefficient 3) is no longer distinguishable from zero at conventional levels (p-value = 0.11), although the estimated effect is indistinguishable from the result obtained in the baseline model without state fixed effects. When using contemporary and Weimar-era fixed effects, the main results remain unchanged. Overall, the evidence reveals that once we account for spatial heterogeneity in a way that avoids post-treatment bias, the results uncovered in HPT remain unchanged.

The same substantive results are obtained.

Discussion

Our goal with this research note is to contribute to the discussion of how to deal with regional heterogeneity in studies of historical legacies. We discuss the specific challenges that the inclusion of fixed effects may pose in this context, i.e., in work that tries to estimate the impact of historical events. We identify two obstacles that scholars need to overcome in order to avoid post-treatment bias when using fixed effects in studies of historical legacies. The first challenge is theoretical. Informed by theory and qualitative evidence, scholars need to decide what type of confounding they want to correct while recognizing that regional fixed effects capture any source of variation across units. This determines which regional units to control for: historical or current ones. In the context of legacy studies, this choice is crucial given that borders are often redrawn throughout history and capture an amorphous set of heterogeneity that may be directly or indirectly on the causal path of interest. The second challenge entails making the correct modeling choices, e.g., correctly specifying the g-estimator to avoid post-treatment bias. The type of regional units used (pre-treatment or post-treatment) defines how they can be incorporated in the analysis and, in turn, whether the results are biased.

We show that these obstacles are real and consequential using the example of PGZ's criticism of HPT. PGZ failed to overcome both of the obstacles listed above, which led to post-treatment bias in their analysis. Properly introducing regional fixed effects in HPT's original analysis – without inducing post-treatment bias – confirms our original results.

As a general recommendation for future studies, if there are concerns about regional confounding along administrative borders that are justified based on the researcher's background knowledge of the case, we suggest using fixed effects based on borders established pre-treatment given the amorphous nature of factors captured by regional fixed effects. However, two other methods are better equipped to deal with geographical heterogeneity and treatments that have localized effects: (1) subsetting the analysis into small (and varying) radii around the source of effects; and (2) sensitivity analyses to spatial autocorrelation. The former solution is already adopted in HPT. We perform the second method in SM 6 and show that spatial correlation is not a relevant threat to inference in this specific context. We believe these approaches offer a more principled solution than using fixed effects to deal with spatial auto-correlation because they reduce researcher degrees of freedom and allow scholars to move beyond arbitrary administrative borders (see Fouka and Voth (2023) for a similar approach).

Our note offers practical reminders about the problem of post-treatment bias and guidance on how to avoid it in historical analyses. We highlight that, while commonly used, the choice of whether or not to include fixed effects is not straightforward in this context. It pays to pause and think whether fixed effects are warranted at all, and if yes, how to properly include them without introducing further bias into the analysis. Ultimately, we hope to highlight the important interplay of theory and empirics, especially in the inherently complicated assessment of the present-day consequences of events that took place decades ago.

Declaration on human subjects

The authors affirm this research did not involve human subjects.

Declaration on conflict of interests

The authors declare no ethical issues or conflicts of interest in this research.

Declaration on data transparency

Research documentation and/or data that support the findings of this study are openly available in the APSR Dataverse at [DOI].

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016a. "The political legacy of American slavery." *The Journal of Politics* 78: 621-641.

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016b. "Explaining causal findings without bias: Detecting and assessing direct effects." *American Political Science Review* 110: 512-529.

Charnysh, Volha, and Evgeny Finkel. 2017. "The Death Camp Eldorado: Political and Economic Effects of Mass Violence." *American Political Science Review* 111: 801-818.

Charnysh, Volha, Eugene Finkel, and Scott Gehlbach. 2023. "Historical Political Economy: Past, Present, and Future." *Annual Review of Political Science* 26: 175-191.

Elwert, Felix, and Christopher Winship. 2014. "Endogenous selection bias: The problem of conditioning on a collider variable." *Annual Review of Sociology* 40: 31-53.

Fouka, Vasiliki, and Hans-Joachim Voth. 2023. "Collective Remembrance and Private Choice: German–Greek Conflict and Behavior in Times of Crisis." *American Political Science Review* 117: 851-870.

Funke, Manuel, Moritz Schularick, and Christoph Trebesch. 2016. "Going to extremes: Politics after financial crises, 1870–2014." *European Economic Review* 88: 227-260.

Hoerner, Julian M., Alexander Jaax, and Toni Rodon. 2019. "The Long-Term Impact of the Location of Concentration Camps on Radical-Right Voting in Germany." *Research & Politics*. DOI: `10.1177/2053168019891376`

Homola, Jonathan, Miguel M. Pereira, and Margit Tavits. 2020. "Legacies of the Third Reich: Concentration Camps and Out-Group Intolerance." *American Political Science Review* 114: 573-590.

Kelly, Morgan. 2019. "The Standard Errors of Persistence." June 3. *Available at SSRN*. DOI: `10.2139/ssrn.3398303`

Kriesi, Hanspeter, Edgar Grande, Romain Lachat, Martin Dolezal, Simon Bornschier, and Timotheos Frey. 2006. "Globalization and the transformation of the national political space: Six European countries compared." *European Journal of Political Research* 45: 921-956.

Lupu, Noam, and Leonid Peisakhin. 2017. "The legacy of political violence across generations." *American Journal of Political Science* 61: 836-851.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How controlling for post-treatment variables can ruin your experiment and what to do about it." *American Journal of Political Science* 62: 760-775.

Norris, Pippa, and Ronald Inglehart. 2018. *Cultural Backlash: The Rise of Authoritarian Populism.* New York: Cambridge University Press.

Pepinsky, Thomas B., Sara Wallace Goodman, and Conrad Ziller. 2023. "Modeling Spatial Heterogeneity and Historical Persistence: Nazi Concentration Camps and Contemporary Intolerance." *American Political Science Review.* DOI: 10.1017/S0003055423000072

Rathenow, Hanns-Fred, and Norbert H. Weber. 1995. "Gedenkstättenbesuche im historisch-politischen Unterricht." In *Praxis der Gedenkstättenpädagogik*, eds. A. Ehmann, W. Kaiser, T. Lutz, H-F. Rathenow, C. vom Stein, N. H. Weber. Opladen: VS Verlag für Sozialwissenschaften: 12-36.

Robins, James M., Donald Blevins, Grant Ritter, and Michael Wulfsohn. 1992. "G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients." *Epidemiology* 3: 319-336.

Simpser, Alberto, Dan Slater, and Jason Wittenberg. 2018. "Dead But Not Gone: Contemporary Legacies of Communism, Imperialism, and Authoritarianism." *Annual Review of Political Science* 21: 419-439.

Weber, Reinhold, and Iris Häuser. 2008. *Baden-Württemberg: Eine kleine politische Landeskunde.* Stuttgart: Landeszentrale für politische Bildung Baden-Württemberg.

Ward, Dalston G. 2019. "Public attitudes toward young immigrant men." *American Political Science Review* 113: 264-269.

Supplementary Materials

"Fixed Effects and Post-Treatment Bias in Legacy Studies"

Jonathan Homola, Miguel M. Pereira, Margit Tavits

November 2023

This supplementary materials file includes the following sections:

SM1: Comparison of model specifications

Together, HPT, PGZ, and the current note provide an extensive list of models with different data sources and model specifications. To help the reader follow this collective effort, Table SM1.1 summarizes the different main specifications modeling the effect of camp proximity on contemporary outcomes.

More specifically, we report the results of the main specifications in HPT in Columns 1 (OLS) and 2 (g-estimator). The main results in PGZ are reported in Columns 3 (contemporary state fixed effects) and 4 (Weimar-era states combined with Prussian provinces fixed effects). Finally, the main results of the new model specifications in the current paper are presented across Columns 5 (contemporary state fixed effects only in the first stage of the g-estimator), 6 (Weimar-era state fixed effects in both stages), and 7 (contemporary state fixed effects only in the first stage *and* Weimar-era state fixed effects in both stages). The models with hybrid fixed effects (Weimar states and Prussian provinces) are discussed in SM 5.

Estimates highlighted in green are in line with the theoretical expectations in HPT and reliable at conventional levels. Estimates highlighted in yellow are in line with the theoretical expectations in HPT but not reliable at conventional levels. Across all the different specifications, the only results that go against the original expectations in HPT can be found in Column 3, which present PGZ's models that likely suffer from post-treatment bias as explained in the main text.

Below we describe the estimation equations for the main models reported in the manuscript. The first equation describes a simple OLS model which only contains pre-treatment variables and where i indexes individuals and j indexes states.

$$\text{Exclusionary Attitudes}_{i,j} = \tau_j + \alpha\text{Distance to Camp} + \theta\mathbf{X'}_{i,j} + \epsilon_{i,j}$$

The vector of Weimar-era state fixed effects (τ_j) captures pre-treatment heterogeneity across states at the time of the creation of the camps. The specification also includes a vector of covariates measured pre-treatment that capture pre-existing political attitudes toward out-groups and local economic conditions ($\mathbf{X'}_{i,j}$). The main parameter of interest is α.

The sequential g-estimator, in turn, allows us to consider different contemporary mediators in addition to the pre-treatment variables. The method starts by estimating a model with both pre-treatment and post-treatment covariates in the first stage:

$$\textbf{Stage 1}: \text{Exclusionary Attitudes}_{i,j} = \tau_j + \alpha\text{Distance to Camp} + \theta\mathbf{X'}_{i,j} + \gamma\mathbf{M'}_{i,j} + \epsilon_{i,j}$$

This equation is similar to the OLS specification described above with the addition of the term **M'**, a vector of post-treatment mediators. It is important to note that the purpose of the first stage is to inform the correction that occurs in the second stage. As such, the coefficients for the post-treatment variables should not be interpreted on their own in a substantive fashion.

Next, the estimator recalculates the outcome variable by removing from it the effects of the mediating variables of interest. In the second stage of the g-estimator, we then regress this "demediated" outcome (*Exclusionary Attitudes'*) on the treatment and pre-treatment covariates as follows:

$$\textbf{Stage 2}: \text{Exclusionary Attitudes'}_{i,j} = \tau_j + \alpha\text{Distance to Camp} + \theta\mathbf{X'}_{i,j} + \epsilon_{i,j}$$

Finally, the results from the sequential g-estimator include bootstrapped standard errors to account for the added uncertainty of its two-step nature.

Table SM1.1: Comparing different model specifications for the effect of camp proximity on contemporary outcomes

	HPT 2020		PGZ 2023		HPT 2023		
	OLS	g-estimator	Contemp. FE	Weimar States + Prussian Prov. FE	Contemp. FE	Weimar FE	Contemp. + Weimar FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
EVS							
Out-group intolerance	-0.011 (0.003)** [N = 2,075]	-0.017 (0.004)** [N = 1,376]	0.005 (0.005) [N = 1,376]	0.001 (0.005) [N = 1,374]	-0.016 (0.004)** [N = 1,376]	-0.010 (0.004)* [N = 2,055]	-0.018 (0.005)** [N = 1,375]
Immigrant resentment	-0.116 (0.017)** [N = 2,075]	-0.106 (0.020)** [N = 1,376]	-0.018 (0.028) [N = 1,376]	-0.043 (0.030) [N = 1,374]	-0.127 (0.022)** [N = 1,376]	-0.051 (0.023)* [N = 2,055]	-0.097 (0.026)** [N = 1,375]
Far-right support	-0.001 (0.001)* [N = 2,075]	-0.003 (0.001)** [N = 1,376]	-0.001 (0.001) [N = 1,376]	-0.001 (0.001) [N = 1,374]	-0.002 (0.001)* [N = 1,376]	-0.002 (0.001) [N = 2,055]	-0.003 (0.001)* [N = 1,375]
Electoral data							
AfD, full sample		-0.081 (0.016)** [N = 10,755]	0.036 (0.011)** [N = 10,870]	see Figure SM5.1**†	-0.058 (0.015)** [N = 10,755]	-0.087 (0.013)** [N = 10,737]	-0.089 (0.014)** [N = 10,737]
AfD, <70km		-0.159 (0.062)** [N = 3,949]		see Figure SM5.1**†	-0.206 (0.059)** [N = 3,949]	-0.220 (0.045)** [N = 3,945]	-0.038 (0.049) [N = 3,945]
AfD+NPD, full sample		-0.092 (0.017)** [N = 10,755]	0.038 (0.011)** [N = 10,870]	see Figure SM5.1**†	-0.067 (0.016)** [N = 10,755]	-0.094 (0.014)** [N = 10,737]	-0.095 (0.015)** [N = 10,737]
AfD+NPD, <70km		-0.171 (0.066)** [N = 3,949]		see Figure SM5.1**†	-0.220 (0.064)** [N = 3,949]	-0.232 (0.048)** [N = 3,945]	-0.036 (0.052) [N = 3,945]
ALLBUS							
Intolerance (foreigners)	-0.030 (0.013)* [N = 3,081]	-0.047 (0.014)** [N = 2,959]					
Intolerance (Jews)	-0.021 (0.009)* [N = 2,886]	-0.029 (0.010)** [N = 2,787]					
Intolerance (Muslims)	-0.026 (0.012)* [N = 3,233]	-0.041 (0.012)** [N = 3,093]					

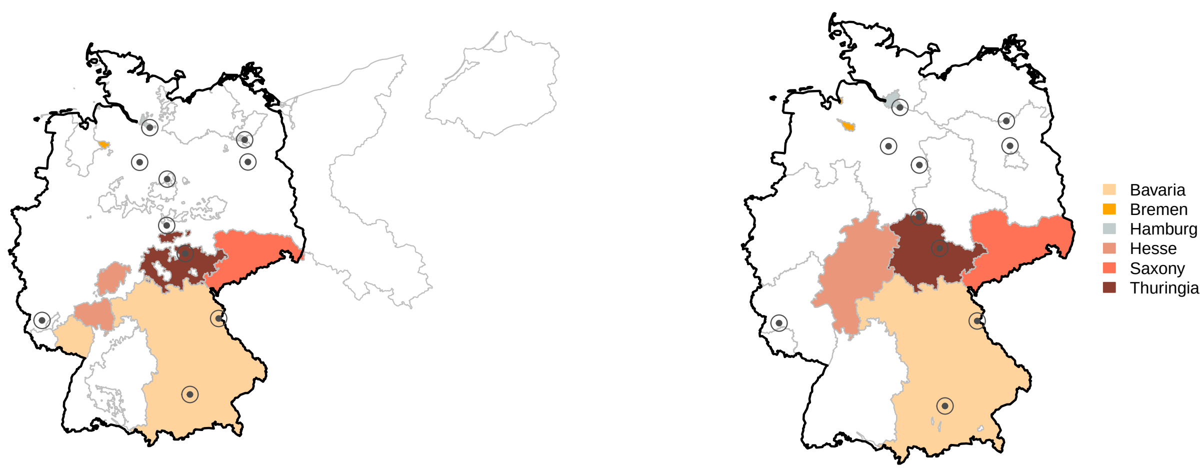
Note: Each row corresponds to an outcome variable, described in the first column. Columns 1 and 2 are based on HPT (2020), Tables 2, 3, and 4. Column 3 refers to PGZ, Table 1, Model 6 for the EVS results and Table 2, Model 2 for the electoral results. Column 4 refers to PGZ, Table A4, Model 6. Columns 5, 6, and 7 are based on the results presented in the main text and in SM 3 and SM 4. In the electoral data analysis, PGZ do not replicate our results using Weimar FE or subsetting the data to specific radii around the camps. Cells highlighted in green indicate estimates that are in line with our expectations and reliable at conventional levels. Cells highlighted in yellow indicate estimates that are in line with our expectations but not reliable at conventional levels. *p<0.05; **p<0.01; † Results not reported in PGZ.

SM2: Contemporary and Weimar-era states

In Figure SM2.1 we compare the composition of German states in the Weimar era and today. Of the 16 German states that exist today, only six existed under the same name before the first concentration camp was built. And even these six regions changed considerably: for two of them (Hamburg and Hesse), less than 50% of their current territory overlaps with their Weimar-era territory,[1] and only Bavaria remains largely unchanged (although the contemporary state no longer includes the Rhenish Palatinate region).[2] In Western Germany, the collection of states as we know them today has only existed since 1957 when the Saar Protectorate rejoined the Federal Republic as the Saarland. In the GDR, in turn, the states were abolished and only reinstated with new borders in 1990 upon the reunification.

Hence, multiple years – in most cases decades – passed between the construction of the camps and the creation of the German states we know today. Any regional heterogeneity captured by contemporary state-level fixed effects is measured post-treatment and is likely to induce post-treatment bias absent a set of assumptions identified by PGZ.

Figure SM2.1: State borders in the Weimar Republic (1932) and in contemporary Germany



(a) Weimar Republic states **(b)** Contemporary states

Note: Panels (a) and (b) describe the state borders in 1932, the year before the first German concentration camp was created, and in contemporary Germany, respectively. In each panel, gray lines indicate the state borders and the dark line corresponds to the current border of Germany. The shaded states in each panel correspond to the six states from the Weimar period that still exist today under the same name.

[1] In addition, Weimar-era Hesse did not include Frankfurt am Main, which is the state's largest city today.

[2] The proportion of territory in these six states overlapping with their Weimar-era counterparts is: 99.2% (Bavaria), 64.5% (Bremen), 39.0% (Hamburg), 29.7% (Hesse), 79.2% (Saxony), and 70.0% (Thuringia). Additionally, note that once Saxony and Thuringia became part of the GDR, they were abolished and divided up into districts (*Bezirke*) in 1952. This example illustrates the fluidity of regional borders over this period.

SM3: Additional analyses using electoral data

This section presents the full results for the main analysis of the electoral data presented in the manuscript, as well as some additional analyses. More specifically, Tables SM3.1-SM3.3 display the full regression results for the analysis presented in Figure 4.

Matching contemporary geographical units with Weimar-era states is not straightforward. First, contemporary districts are not always contained within a single Weimar-era state. Second, the state of Saarland was not part of Weimar Germany. We therefore use three alternative methods to interpolate Weimar states: (1) matching each contemporary district to the Weimar state that overlaps with the district's geographical center (*centroid interpolation*), (2) using the same centroid interpolation but including Saarland respondents/districts as an additional Weimar state, and (3) matching each contemporary district to the Weimar state that overlaps with the largest share of its area (*area interpolation*). The analyses reported in the main text only included the first interpolation method. However, Figure SM3.1 shows that the findings are also robust to using the other matching approaches described (i.e., centroid interpolation while also including Saarland respondents, and area interpolation).

Table SM3.1: The controlled direct effect of camp proximity on support for radical right parties in 2017, with contemporary state fixed effects in first stage of g-estimator

	AfD Vote Share		AfD + NPD Vote Share	
	Full sample	$< 70km$	Full sample	$< 70km$
	(1)	(2)	(3)	(4)
Distance	-0.058^{**}	-0.206^{**}	-0.067^{**}	-0.220^{**}
(in 10kms)	(0.015)	(0.059)	(0.016)	(0.064)
Nazi party share (1933)	0.008	-0.004	0.012^{*}	0.0004
	(0.005)	(0.010)	(0.006)	(0.010)
% Unemployed (1933)	0.168^{**}	0.120^{**}	0.177^{**}	0.129^{**}
	(0.028)	(0.033)	(0.030)	(0.036)
Population (1925)	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00001)	(0.00000)	(0.00001)
% Jews (1925)	-2.969^{**}	-4.306^{**}	-3.144^{**}	-4.604^{**}
	(0.147)	(0.458)	(0.155)	(0.484)
Current state FEs (N=16)	✓	✓	✓	✓
Contemporary variables	✓	✓	✓	✓
Observations	10,755	3,949	10,755	3,949
Adjusted R^2	0.058	0.084	0.059	0.086

Note: Entries are coefficients of the controlled direct effect of distance to closest camp on support for the AfD (Columns 1-2) and AfD+NPD (Column 3-4) in 2017, corresponding to Table 4 in HPT. All models report the second stage of the sequential g-estimator (bootstrapped standard errors in parentheses). All models include contemporary state fixed effects and contemporary mediators and confounders in the *first* stage regression. *p<0.05; **p<0.01

Table SM3.2: The controlled direct effect of camp proximity on support for radical right parties in 2017, accounting for systematic differences across Weimar states (interpolated from centroids of contemporary Gemeinden)

	AfD Vote Share		AfD + NPD Vote Share	
	Full sample	$< 70km$	Full sample	$< 70km$
	(1)	(2)	(3)	(4)
Distance	-0.087^{**}	-0.220^{**}	-0.094^{**}	-0.232^{**}
(in 10kms)	(0.013)	(0.045)	(0.014)	(0.048)
Nazi party share (1933)	-0.028^{**}	-0.039^{**}	-0.028^{**}	-0.040^{**}
	(0.004)	(0.008)	(0.004)	(0.008)
% Unemployed (1933)	0.055^{**}	0.027	0.054^{**}	0.028
	(0.015)	(0.019)	(0.016)	(0.021)
Population (1925)	−0.00000	0.00001	−0.00000	0.00001
	(0.00000)	(0.00001)	(0.00000)	(0.00001)
% Jews (1925)	-0.505^{**}	-1.001^{**}	-0.519^{**}	-1.105^{**}
	(0.117)	(0.260)	(0.122)	(0.272)
Weimar state FEs (N=17)	✓	✓	✓	✓
Contemporary variables	✓	✓	✓	✓
Observations	10,737	3,945	10,737	3,945
Adjusted R^2	0.390	0.361	0.389	0.364

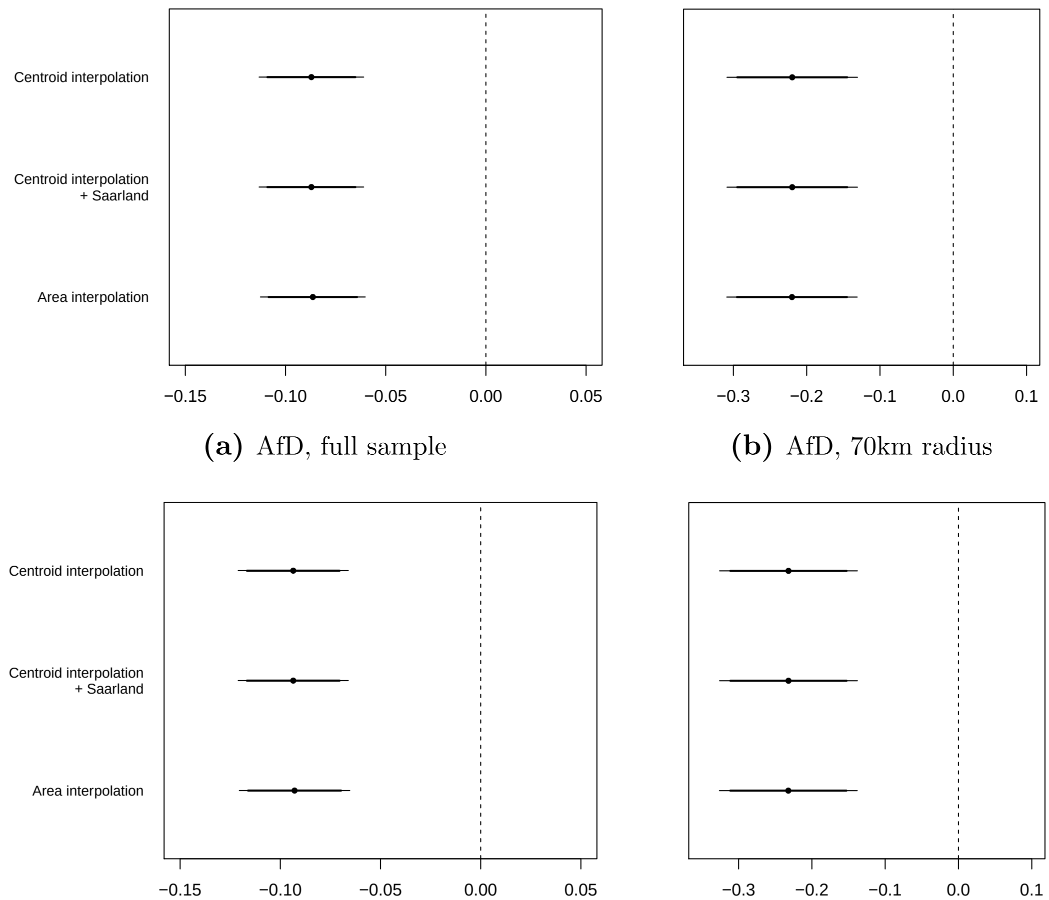
Note: Entries are coefficients of the controlled direct effect of distance to closest camp on support for the AfD (Columns 1-2) and AfD+NPD (Column 3-4) in 2017, corresponding to Table 4 in HPT. All models report the second stage of the sequential g-estimator (bootstrapped standard errors in parentheses). All models include Weimar state fixed effects (interpolated based on the centroids of contemporary Gemeinden) in *both* stages and contemporary mediators and confounders in the *first* stage regression. *p<0.05; **p<0.01

Table SM3.3: The controlled direct effect of camp proximity on support for radical right parties in 2017, with current state fixed effects in first stage of g-estimator and accounting for systematic differences across Weimar states (interpolated from centroids of contemporary Gemeinden)

	AfD Vote Share		AfD + NPD Vote Share	
	Full sample	$< 70km$	Full sample	$< 70km$
	(1)	(2)	(3)	(4)
Distance	-0.089^{**}	-0.038	-0.095^{**}	-0.036
(in 10kms)	(0.014)	(0.049)	(0.015)	(0.052)
Nazi party share (1933)	0.009^{*}	0.006	0.012^{**}	0.009
	(0.004)	(0.009)	(0.005)	(0.010)
% Unemployed (1933)	0.080^{**}	0.085^{*}	0.081^{**}	0.091^{*}
	(0.018)	(0.033)	(0.019)	(0.036)
Population (1925)	-0.00000	0.00000	-0.00000^{*}	0.00000
	(0.00000)	(0.00001)	(0.00000)	(0.00001)
% Jews (1925)	-1.088^{**}	-1.914^{**}	-1.150^{**}	-2.089^{**}
	(0.127)	(0.354)	(0.132)	(0.374)
Current state FEs (N=16)	✓	✓	✓	✓
Weimar state FEs (N=17)	✓	✓	✓	✓
Contemporary variables	✓	✓	✓	✓
Observations	10,737	3,945	10,737	3,945
Adjusted R^2	0.415	0.410	0.416	0.417

Note: Entries are coefficients of the controlled direct effect of distance to closest camp on support for the AfD (Columns 1-2) and AfD+NPD (Column 3-4) in 2017, corresponding to Table 4 in HPT. All models report the second stage of the sequential g-estimator (bootstrapped standard errors in parentheses). All models include contemporary state fixed effects and contemporary mediators and confounders in the *first* stage regression. All models also include Weimar state fixed effects (interpolated based on the centroids of contemporary Gemeinden) in *both* stages. *p<0.05; **p<0.01

Figure SM3.1: The controlled direct effect of camp proximity on support for radical right parties in 2017, with alternative methods to interpolate Weimar states from contemporary Gemeinden



(a) AfD, full sample **(b)** AfD, 70km radius

(c) AfD+NPD, full sample **(d)** AfD+NPD, 70km radius

Note: Plots depict estimates and 95/90% confidence intervals from the sequential g-estimator for the controlled direct effects of distance to camps on support for radical right parties in 2017 (described in each panel label). Each estimate corresponds to a different model based on alternative methods to match contemporary districts with Weimar-era states.

SM4: Additional analyses using EVS data

This section presents supplementary analyses of the EVS data. More specifically, Tables SM4.1-SM4.3 display the full regression results for the analysis presented in Figure 5. In the main text, we only report results with *centroid interpolation*. However, Figure SM4.1 shows that the findings are also robust to using the other matching approaches described (i.e., centroid interpolation while also including Saarland respondents and area interpolation). Finally, while these main models are all replications of our g-estimation approach, the results in Figure SM4.2 show that the "pre-treatment only" OLS models are also robust to the properly specified inclusion of state fixed effects.

Table SM4.1: The controlled direct effect of camp proximity on contemporary attitudes, with current state fixed effects in first stage of g-estimator

	Outgroup Intolerance		Immigrant Resentment		Support Far-Right Parties	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to camp	0.005	−0.016**	−0.023	−0.127**	−0.0003	−0.002*
(in 10kms)	(0.005)	(0.004)	(0.029)	(0.022)	(0.001)	(0.001)
% Jews (1925)	7.561	−0.782	−5.063	0.185	1.891	0.336
	(6.289)	(1.609)	(35.980)	(11.344)	(1.564)	(0.454)
% Unemployed (1933)	1.321	2.177*	1.837	12.369*	0.266	0.299
	(0.909)	(0.934)	(5.202)	(5.694)	(0.226)	(0.195)
Population (1925)	−0.017	−0.013	−0.108	−0.104	−0.006	−0.0004
	(0.016)	(0.012)	(0.093)	(0.077)	(0.004)	(0.003)
Nazi party share (1933)	−0.726**	−0.244	−5.109**	−3.705*	−0.170*	−0.078
	(0.280)	(0.242)	(1.605)	(1.696)	(0.070)	(0.063)
Contemporary covariates						
Conservatism	0.049**		0.292**		0.039**	
	(0.009)		(0.054)		(0.002)	
Unemployed	0.026		0.971**		0.095**	
	(0.054)		(0.309)		(0.013)	
Education	−0.092**		−0.642**		−0.014**	
	(0.013)		(0.075)		(0.003)	
Female	−0.133**		−0.577**		−0.033**	
	(0.036)		(0.204)		(0.009)	
Age	0.002		0.029**		−0.001**	
	(0.001)		(0.006)		(0.0003)	
% Immigrants (2007)	−1.434		−10.017*		−0.208	
	(0.748)		(4.282)		(0.186)	
% Unemployed (2007)	−2.472**		−17.881**		0.244	
	(0.740)		(4.232)		(0.184)	
Urban	0.022		0.041		0.001	
Model	G-est. Stage 1	G-est. Stage 2	G-est. Stage 1	G-est. Stage 2	G-est. Stage 1	G-est. Stage 2
Current state FEs (N=16)	✓		✓		✓	
Observations	1,376	1,376	1,376	1,376	1,376	1,376
Adjusted R^2	0.117	0.025	0.215	0.042	0.240	0.016

Note: Entries are coefficients of the effect of distance to closest camp on different outcomes, described in the column headers. Model 1, 3, and 5 correspond to the first stage of the sequential g-estimation (standard errors in parentheses), with contemporary covariates including current state fixed effects. Models 2, 4, and 6, represent the second stage in the sequential g-estimation (bootstrapped standard errors in parentheses). *p<0.05; **p<0.01

Table SM4.2: Effects of camp proximity on out-group intolerance, immigrant resentment, and support for far-right parties (EVS), accounting for systematic differences across Weimar states (interpolated from centroids of contemporary Kreise)

	Out-group Intolerance		Immigrant Resentment		Support Far-Right Parties	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to camp	−0.011**	−0.010*	−0.077**	−0.051*	−0.001	−0.002
(in 10kms)	(0.003)	(0.004)	(0.020)	(0.023)	(0.001)	(0.001)
% Jews (1925)	−1.876	−0.580	8.532	16.023	−0.077	0.183
	(1.192)	(1.494)	(7.037)	(9.978)	(0.268)	(0.474)
% Unemployed (1933)	2.344**	3.128**	0.143	11.169	0.107	0.682**
	(0.719)	(0.915)	(4.243)	(5.799)	(0.162)	(0.237)
Population (1925)	−0.037**	0.005	−0.300**	−0.064	−0.003	−0.001
	(0.012)	(0.014)	(0.071)	(0.099)	(0.003)	(0.003)
Nazi party share (1933)	−0.466*	−0.418	−2.796*	−5.806**	−0.021	−0.082
	(0.196)	(0.217)	(1.156)	(1.490)	(0.044)	(0.056)
Model	OLS	G-est. Stage 2	OLS	G-est. Stage 2	OLS	G-est. Stage 2
Weimar state FEs (N = 16)	✓	✓	✓	✓	✓	✓
Contemporary variables		✓		✓		✓
Observations	2,055	1,375	2,055	1,375	2,055	1,375
Adjusted R^2	0.045	0.057	0.075	0.096	0.005	0.034

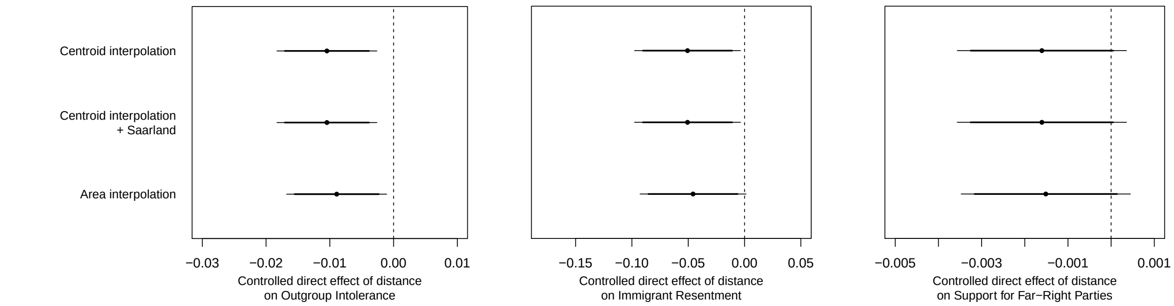
Note: Entries are estimates of the effect of distance to closest camp on the different outcomes, described in column headers. Models 1, 3, and 5 account exclusively for interwar covariates (standard errors in parentheses). Models 2, 4, and 6 are the 2nd stage of the sequential g-estimator to also account for contemporary predictors (bootstrapped standard errors in parentheses). All models (and both stages of the g-estimator) include Weimar state fixed effects (interpolated based on the centroids of contemporary Kreise). *p<0.05; **p<0.01

Table SM4.3: The controlled direct effect of camp proximity on out-group intolerance, immigrant resentment, and support for far-right parties (EVS), with current state fixed effects in first stage of g-estimator and accounting for systematic differences across Weimar states (interpolated from centroids of contemporary Kreise)

	Outgroup Intolerance		Immigrant Resentment		Support Far-Right Parties	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to camp	0.005	-0.018^{**}	-0.021	-0.097^{**}	-0.0005	-0.003^{*}
	(0.005)	(0.005)	(0.030)	(0.027)	(0.001)	(0.001)
% Jews (1925)	10.570	-1.690	26.777	9.171	1.462	0.002
	(6.733)	(1.486)	(38.371)	(9.552)	(1.671)	(0.461)
% Unemployed (1933)	2.329^{*}	3.142^{**}	2.808	7.614	0.315	0.604^{*}
	(1.088)	(0.971)	(6.200)	(6.010)	(0.270)	(0.247)
Population (1925)	-0.024	-0.007	-0.144	-0.173	-0.003	-0.003
	(0.017)	(0.016)	(0.098)	(0.111)	(0.004)	(0.004)
Nazi party share (1933)	-0.650^{*}	-0.249	-4.861^{**}	-4.632^{**}	-0.166^{*}	-0.055
	(0.288)	(0.229)	(1.642)	(1.636)	(0.071)	(0.060)
Contemporary covariates						
Conservatism	0.049^{**}		0.297^{**}		0.039^{**}	
	(0.009)		(0.054)		(0.002)	
Unemployed	0.026		0.933^{**}		0.094^{**}	
	(0.054)		(0.309)		(0.013)	
Education	-0.092^{**}		-0.647^{**}		-0.014^{**}	
	(0.013)		(0.075)		(0.003)	
Female	-0.133^{**}		-0.559^{**}		-0.034^{**}	
	(0.036)		(0.204)		(0.009)	
Age	0.001		0.028^{**}		-0.001^{**}	
	(0.001)		(0.006)		(0.0003)	
% Immigrants (2007)	-1.773^{*}		-12.225^{**}		-0.296	
	(0.790)		(4.503)		(0.196)	
% Unemployed (2007)	-2.599^{**}		-19.576^{**}		0.085	
	(0.768)		(4.380)		(0.191)	
Urban	0.019		0.073		0.003	
	(0.013)		(0.073)		(0.003)	
Model	G-est. Stage 1	G-est. Stage 2	G-est. Stage 1	G-est. Stage 2	G-est. Stage 1	G-est. Stage 2
Current state FEs (N=16)	✓		✓		✓	
Weimar state FEs (N=16)	✓	✓	✓	✓	✓	✓
Observations	1,375	1,375	1,375	1,375	1,375	1,375
Adjusted R^2	0.118	0.039	0.221	0.080	0.245	0.019

Note: Entries are coefficients of the effect of distance to closest camp on different outcomes, described in the column headers. Model 1, 3, and 5 correspond to the first stage of the sequential g-estimation (standard errors in parentheses), with contemporary covariates including current state fixed effects. Models 2, 4, and 6, represent the second stage in the sequential g-estimation (bootstrapped standard errors in parentheses). All models also include Weimar state fixed effects (interpolated based on the centroids of contemporary Gemeinden) in *both* stages. *p<0.05; **p<0.01

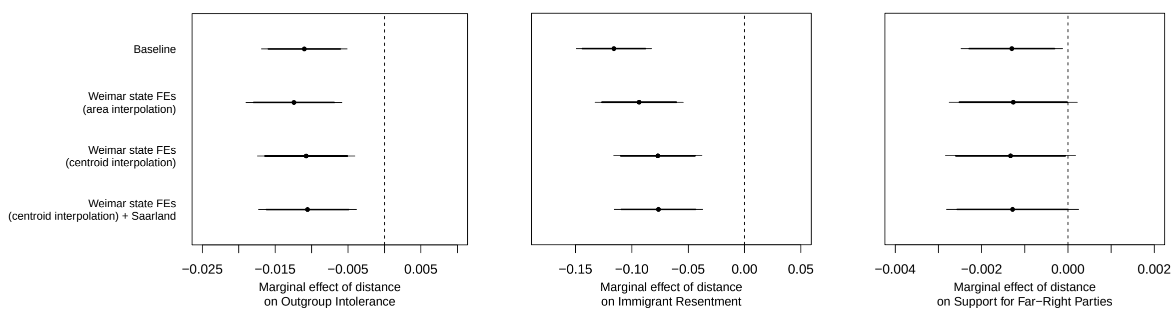
Figure SM4.1: The controlled direct effect of camp proximity on outgroup intolerance, immigrant resentment, and support for far-right parties (EVS), with alternative methods to interpolate Weimar states from contemporary Kreise



(a) Out-group Intolerance **(b)** Immigrant Resentment **(c)** Far-Right Support

Note: Plots depict estimates and 95/90% confidence intervals from the sequential g-estimator for the controlled direct effects of distance to camps on contemporary attitudes (described in each panel). Each estimate corresponds to a different model based on alternative methods to match contemporary districts with Weimar-era states. Top estimates in each model come from models 2, 4, and 6 in Table SM4.2. Due to space constraints,the models estimated to obtain the remaining estimates are reported in the replication documents on Harvard Dataverse.

Figure SM4.2: Effects of camp proximity on out-group intolerance, immigrant resentment, and support for far-right parties (EVS), accounting for state-level heterogeneity



(a) Out-group Intolerance **(b)** Immigrant Resentment **(c)** Far-Right Support

Note: Plots depict estimates and 95/90% confidence intervals from OLS models for the effects of distance to camps on contemporary attitudes (described in each panel label). Each estimate derives from a different model specification, described on the y-axis. The baseline specification corresponds to the results reported in Table 2 (Models 1, 3, and 5) in HPT. Centroid interpolation estimates come from models 1, 3, and 5 in Table SM4.2. Due to space constraints,the models estimated to obtain the remaining estimates are reported in the replication documents on Harvard Dataverse.

SM5: Prussian provinces: discussion and additional analyses

PGZ's original critique (2020) includes an analysis with Weimar-era state fixed effects. Although the authors use a map from 1925, the results reported are substantively similar to those we report here and inconsistent with PGZ's own argument. They then further split the analyses, treating Prussian internal provinces as separate states, which renders the effects of camp proximity unreliable. In the most recent version of the critique, PGZ only report the models with a combination of Weimar states and Prussian provinces. In this section, we take a closer look at these analyses and point out that the decision to split Prussia into provinces appears arbitrary and atheoretical. We then proceed to show that when we include province-level fixed effects in our electoral analysis, the main findings also remain unchanged – something PGZ did not report.

1. *Arbitrary decision to include Prussian provinces:* PGZ's main argument for the inclusion of Prussian provinces in the analyses with Weimar-era fixed effects is that provinces are the historical antecedents of contemporary states. This argument makes sense if the goal is to identify regions that match the geography of contemporary states as closely as possible. However, the goal of accounting for regional differences in a model is to absorb heterogeneity that (a) is of theoretical interest and (b) results from the socioeconomic and political variation in the regions *before* exposure to treatment. Trying to identify pre-treatment areas based on the shapes of post-treatment areas goes against this idea. Moreover, the decision to include fixed effects for Weimar-era Germany's *states* and Prussia's *provinces* in the same model specification seems arbitrary. Such a setup is equivalent to including state fixed effects for some US states, and county (or congressional district) fixed effects for others. Although California might be more heterogeneous than Rhode Island, we are not aware that this is a common empirical approach.[3]

2. *Atheoretical approach:* PGZ claim that within-Prussia heterogeneity necessitates the inclusion of province fixed effects. As an example, in the original critique the authors discuss a failed Reichsreform which planned to divide Prussia into several sub-states. We actually believe this example makes an argument *against* the inclusion of province fixed effects. The reform failed because Prussia did not want to lose the influence it had over the federal government (as the largest state). It was the opposition of the

[3]Another comparison could be the inclusion of fixed effects for the nine English regions (i.e., East Midlands, East of England, London, North East, North West, South East, South West, West Midlands, and Yorkshire and the Humber) alongside fixed effects for Scotland, Wales, and Northern Ireland in studies focusing on the UK. We are not aware that this is a common approach either.

Prussian state as a whole – not of specific provinces – that blocked this reform (Schulz 1963). As such, the failed reform is an example of Prussian unity, not division. Holborn (1956: 335) explicitly describes that in "Prussia itself [...] no strong signs could be found that the provinces wished to become states." In fact, the Weimar constitution included a provision for the possible secession of individual provinces through plebiscite. As far as we can tell, this instrument was used only once, in Upper Silesia, where over 90% voted to remain part of Prussia in 1922 (Hertz-Eichenrode 1969; Schattkowsky 1994). Another attempt to schedule a plebiscite in Hanover in 1924 failed because there was not enough interest among voters (Funk 2010; Heimann 2011). This is in line with other accounts that emphasize the relevance and strength of Prussia as a whole (Orlow 1991). In other words, these historical accounts provide little evidence to support strong concerns regarding Prussia's own inherent regional heterogeneity.[4] We believe this discussion highlights once again the importance of having well articulated theoretical arguments to motivate one's modeling choices. For instance, if one is interested in capturing differences in school curricula that are determined at the state level, state-level fixed effects should be included. If instead one is interested in capturing administrative differences that vary at the substate level, then the analysis should include fixed effects at the level of the administrative region below the state.[5] Importantly, choosing one level in one state and another in the remaining states is a decision that seems difficult to motivate theoretically.

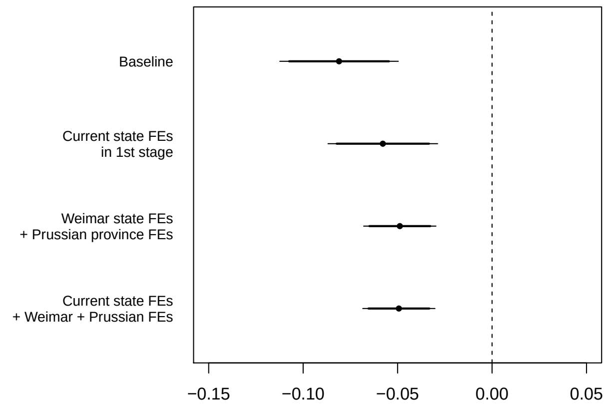
3. *Prussian provinces and electoral data:* Despite our concerns regarding PGZ's arbitrary and atheoretical inclusion of province fixed effects, we decided to replicate our electoral analysis while accounting for Prussia's provinces. More specifically, we follow the analysis in the main text and replicate Table 4 in HPT while including (1) contemporary state-level fixed effects in the first stage of the g-estimator, (2) Weimar-era state-level and Prussian province-level fixed effects in both stages of the g-estimator, and (3) both contemporary states *and* Weimar-era states and Prussian provinces. The results in Figure SM5.1 show that our main conclusions are robust to this hybrid approach of using fixed effects at the province level within Prussia and at the state level for the rest of the country. Across the different specifications, we see that the effect of distance is

[4] This does not mean that all of Prussia was always perfectly united. For example, there was a limited and ultimately unsuccessful independence movement in the Rhine Province that led to the declaration of a short-lived "Rhenish Republic" in the mid-1920s (Epstein 1967).

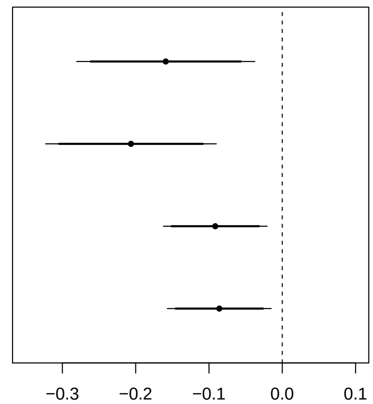
[5] In the case of Weimar Germany, choosing a level below the state is more challenging because of the differences in administrative setups across states. However, all states were ultimately divided up into a combination of Ämter, Kreise, and Regierungsbezirke, which would allow for the inclusion of fixed effects at that level.

always negative and statistically reliable at conventional levels. PGZ neglect to report these results.

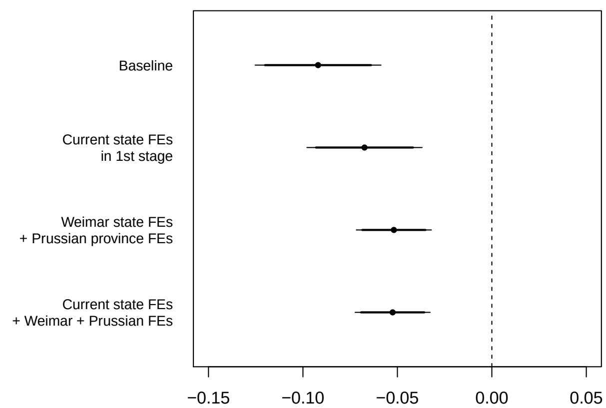
Figure SM5.1: The controlled direct effect of camp proximity on support for radical right parties in 2017, accounting for state-level and Prussian province-level heterogeneity



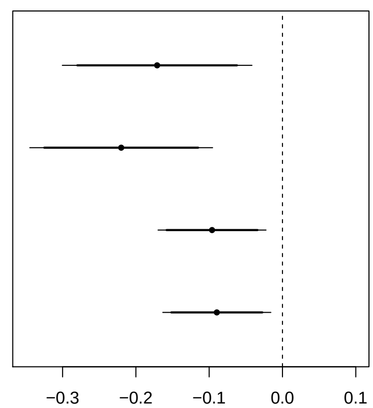
(a) AfD, full sample



(b) AfD, 70km radius



(c) AfD+NPD, full sample



(d) AfD+NPD, 70km radius

Note: Plots depict estimates and 95/90% confidence intervals from the sequential g-estimator for the controlled direct effects of distance to camps on support for radical right parties in 2017 (described in each panel label). Each estimate corresponds to a different model specification, described on the y-axis. The baseline specification corresponds to the results reported in Table 4 HPT.

SM6: Noise simulations

PGZ's critique is motivated in part by growing concerns about the effects of spatial correlation in the historical legacies literature (Kelly 2019). Although state-level heterogeneity – when properly incorporated in the analyses – does not explain the findings in HPT, spatial correlation may still play a role. To directly assess the robustness of findings to spatial correlation, Kelly suggests reestimating the main models with spatially-correlated noise.

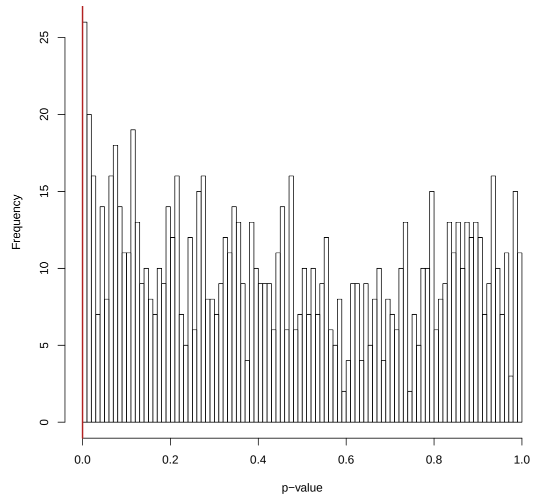
In this section, we conduct noise simulations to investigate the extent to which the main findings in HPT might be explained by spatial noise. Following Fouka and Voth (2023), we replace the geographic variable – *Distance* – with spatially correlated noise. The procedure consists of running 1,000 replications of the main models, each time with a different vector of spatially correlated random noise replacing *Distance*. Following Kelly, this spatial noise is drawn from a multivariate normal distribution using a variance-covariance matrix based on the Matern function. For the Matern function, we set the variance and shape parameters to be 1. For the crucially important correlation range, we follow Kelly's recommendations for the analysis of German data and present our results for a correlation range of 3 degrees and a correlation range of 5 degrees.

For each observation, we use this setup to draw 1,000 iterations of spatial noise. We then run 1,000 regressions replicating our g-estimation models, where the spatial noise replaces the *Distance* variable.[6] For every regression, we store the p-value of the spatial noise variable. The distribution of these p-values is then plotted in Figures SM6.1-SM6.4, along with a red vertical line illustrating the p-values from the original regressions. Across the different datasets and outcome variables, spatial noise very rarely outperforms our *Distance* variable in terms of explanatory power.

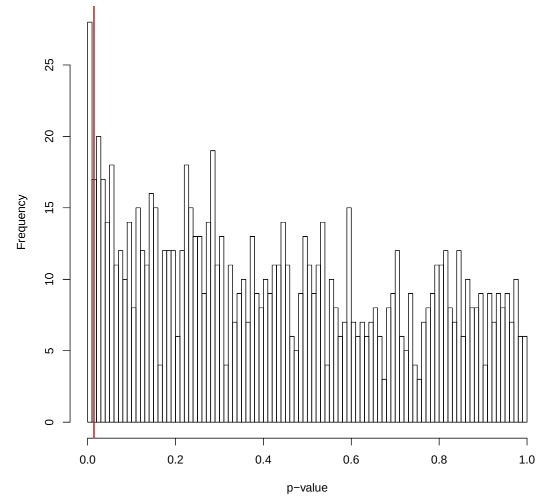
Table SM6.1 summarizes the main results from this simulation exercise. More specifically, it shows the amount of times that spatial noise had more explanatory power than HPT's original treatment variable across the 1,000 simulations. The results suggest that spatially-correlated noise rarely outperforms the *Distance* variable. As a reference point, in most of Kelly's replications of papers based on European data, the explanatory power of spatial noise outperformed the original predictor in 20-50% of all cases. None of HPT's models approaches these numbers. Spatial noise outperformed camp proximity less than 5% of the time in 11 of the 14 sets of simulations performed. The only exceptions are the electoral data models within 70km radii and a 5 degrees correlation range, where noise outperformed the predictor 11% of the time. The analyses suggest that camp proximity captures something more meaningful than mere spatial noise.

[6] Given the computationally intensive nature of this method, we report regular (i.e., non-bootstrapped) standard errors for these models.

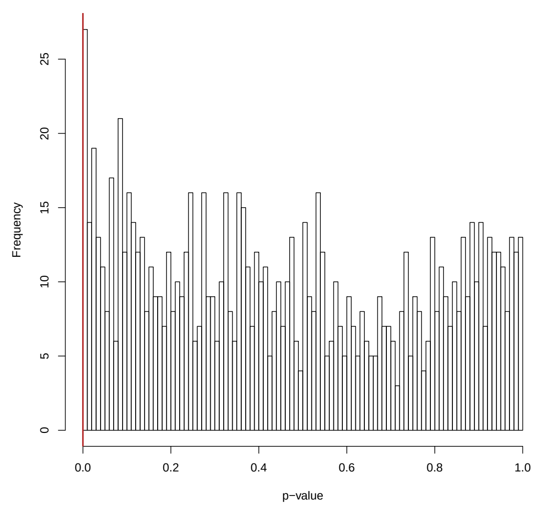
Figure SM6.1: The controlled direct effect of spatial noise on support for radical right parties in 2017 (correlation range of 3 degrees)



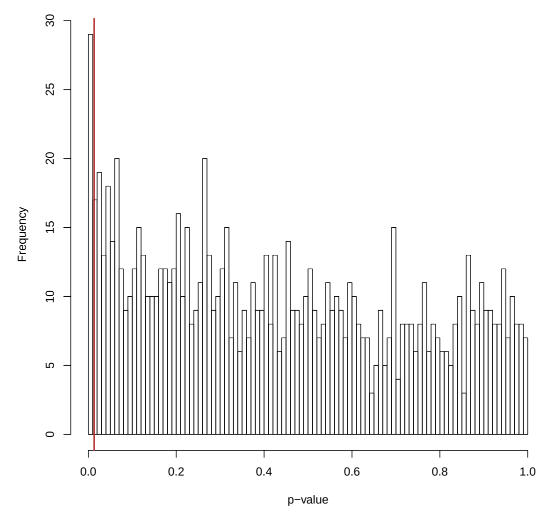
(a) AfD, full sample



(b) AfD, 70km radius



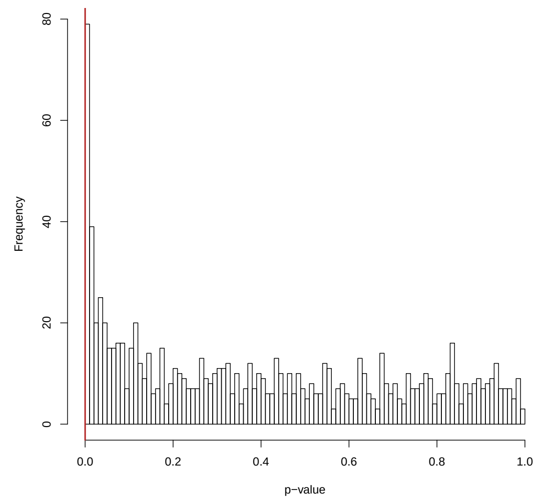
(c) AfD+NPD, full sample



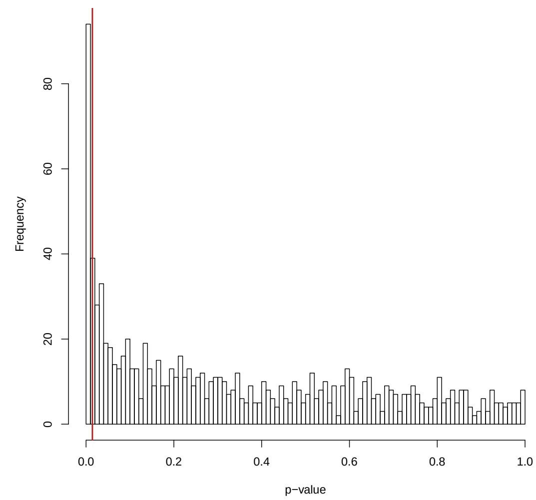
(d) AfD+NPD, 70km radius

Note: Plots show the distribution of p-values resulting from 1,000 simulations of the sequential g-estimator for the controlled direct effects of distance to camps on support for radical right parties in 2017 (described in each panel label). In each iteration, *Distance* has been replaced by simulated spatially correlated noise according to the Matern function, with a variance and shape of 1 and a correlation range of 3 degrees, following Kelly (2019) and Fouka and Voth (2023). The red vertical line indicates the p-value from the original model specification. The explanatory power of spatial noise is higher than that of our *Distance* variable 0.0% (panel a), 3.8% (panel b), 0.0% (panel c), and 3.4% (panel d) of the time respectively.

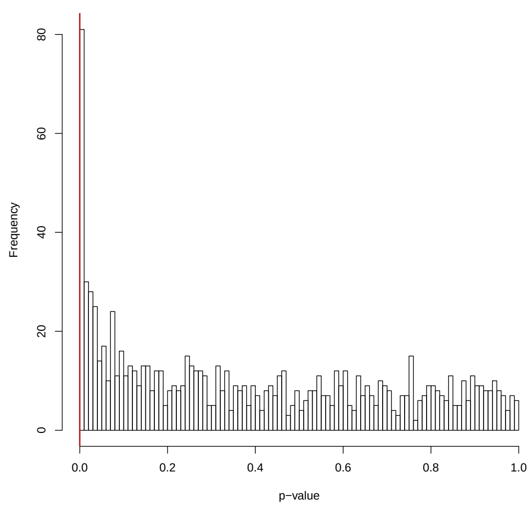
Figure SM6.2: The controlled direct effect of spatial noise on support for radical right parties in 2017 (correlation range of 5 degrees)



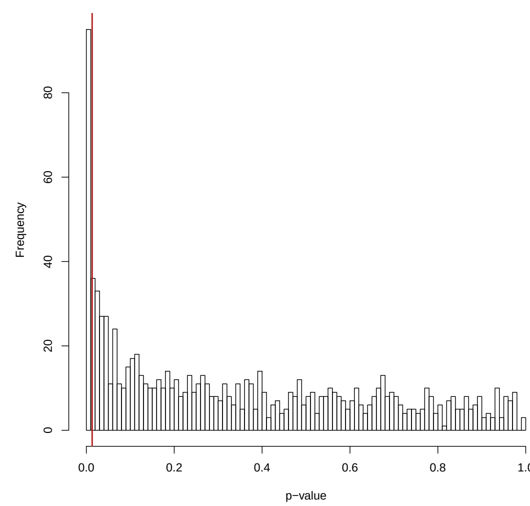
(a) AfD, full sample



(b) AfD, 70km radius



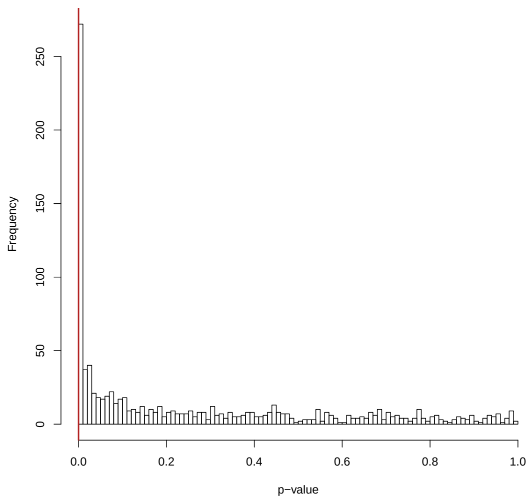
(c) AfD+NPD, full sample



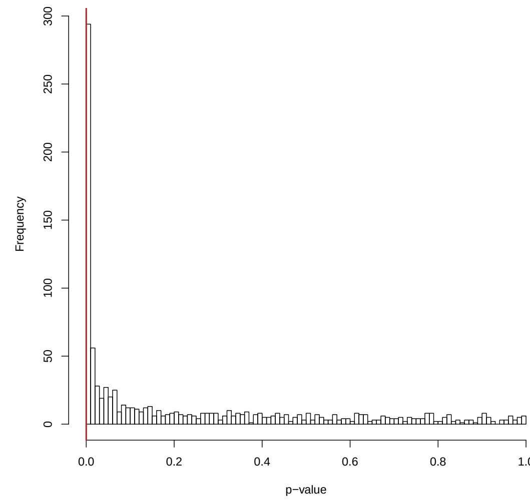
(d) AfD+NPD, 70km radius

Note: Plots show the distribution of p-values resulting from 1,000 simulations of the sequential g-estimator for the controlled direct effects of distance to camps on support for radical right parties in 2017 (described in each panel label). In each iteration, *Distance* has been replaced by simulated spatially correlated noise according to the Matern function, with a variance and shape of 1 and a correlation range of 5 degrees, following Kelly (2019) and Fouka and Voth (2023). The red vertical line indicates the p-value from the original model specification. The explanatory power of spatial noise is higher than that of our *Distance* variable 0.0% (panel a), 11.2% (panel b), 0.0% (panel c), and 11.1% (panel d) of the time respectively.

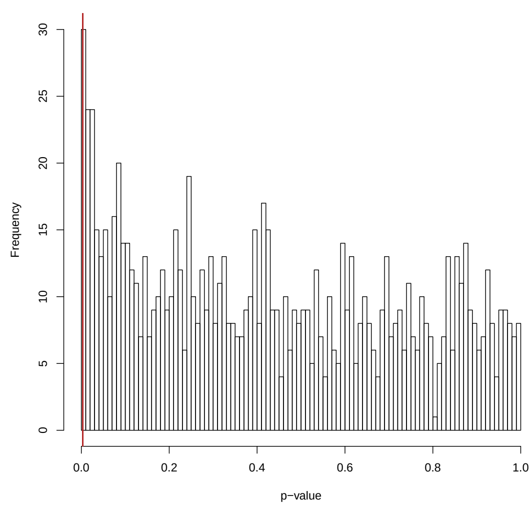
Figure SM6.3: The controlled direct effect of spatial noise on outgroup intolerance, immigrant resentment, and support for far-right parties (correlation range of 3 degrees)



(a) Outgroup Intolerance



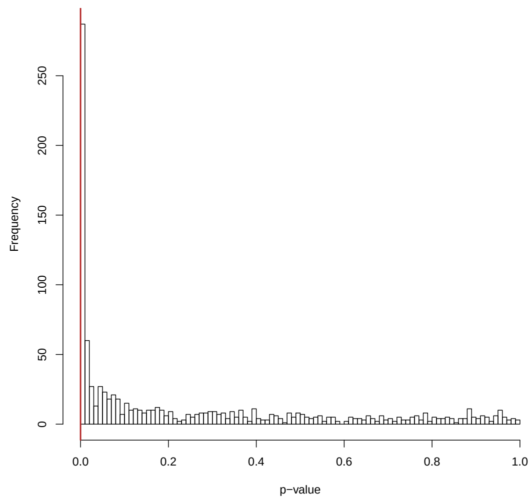
(b) Immigrant Resentment



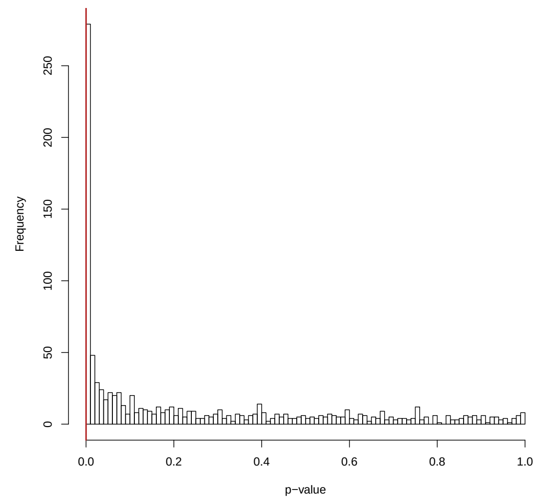
(c) Far-Right Support

Note: Plots show the distribution of p-values resulting from 1,000 simulations of the sequential g-estimator for the controlled direct effects of distance to camps on contemporary attitudes (described in each panel label). In each iteration, *Distance* has been replaced by simulated spatially correlated noise according to the Matern function, with a variance and shape of 1 and a correlation range of 3 degrees, following Kelly (2019) and Fouka and Voth (2023). The red vertical line indicates the p-value from the original model specification. The explanatory power of spatial noise is higher than that of our *Distance* variable 5.1% (panel a), 0.4% (panel b), and 1.1% (panel c) of the time respectively.

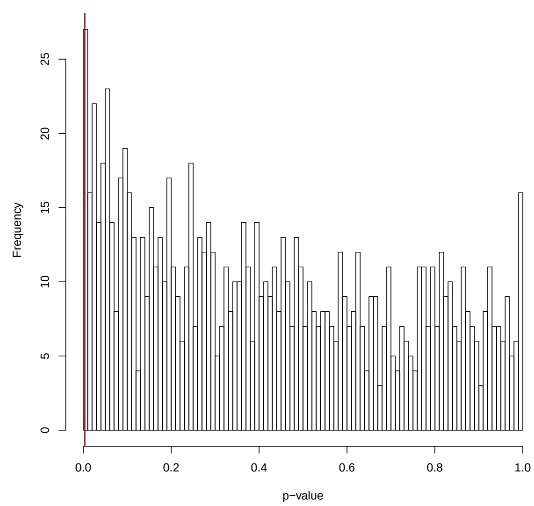
Figure SM6.4: The controlled direct effect of spatial noise on outgroup intolerance, immigrant resentment, and support for far-right parties (correlation range of 5 degrees)



(a) Outgroup Intolerance



(b) Immigrant Resentment



(c) Far-Right Support

Note: Plots show the distribution of p-values resulting from 1,000 simulations of the sequential g-estimator for the controlled direct effects of distance to camps on contemporary attitudes (described in each panel label). In each iteration, *Distance* has been replaced by simulated spatially correlated noise according to the Matern function, with a variance and shape of 1 and a correlation range of 5 degrees, following Kelly (2019) and Fouka and Voth (2023). The red vertical line indicates the p-value from the original model specification. The explanatory power of spatial noise is higher than that of our *Distance* variable 4.5% (panel a), 0.9% (panel b), and 1.1% (panel c) of the time respectively.

Table SM6.1: Explanatory power of the effect of spatial noise on contemporary outcomes

	Correlation range	
	3 degrees	5 degrees
Electoral data		
AfD, full sample	0.0%	0.0%
AfD, 70km radius	3.8%	11.2%
AfD+NPD, full sample	0.0%	0.0%
AfD + NPD, 70km radius	3.4%	11.1%
EVS		
Outgroup intolerance	5.1%	4.5%
Immigrant resentment	0.4%	0.9%
Far-right support	1.1%	1.1%

Note: Entries indicate how often spatial noise outperforms the original distance variable in explaining the different contemporary outcomes. For more details, see SM 6.

SM7: References

Epstein, Klaus. 1967. "Adenauer and Rhenish Separatism." *The Review of Politics* 29: 536-545.

Fouka, Vasiliki, and Hans-Joachim Voth. 2023. "Collective Remembrance and Private Choice: German–Greek Conflict and Behavior in Times of Crisis." *American Political Science Review* 117: 851-870.

Funk, Albert. 2010. *Kleine Geschichte des Föderalismus: vom Fürstenbund zur Bundesrepublik.* Paderborn: Ferdinand Schöningh.

Heimann, Siegfried. 2011. *Der Preußische Landtag 1899-1947: eine politische Geschichte.* Berlin: Links.

Hertz-Eichenrode, Dieter. 1969. *Politik und Landwirtschaft in Ostpreußen 1919-1930: Untersuchung eines Strukturproblems in der Weimarer Republik.* Köln, Opladen: Westdeutscher Verlag.

Holborn, Hajo. 1956. "Prussia and the Weimar Republic." *Social Research* 23: 331-342.

Kelly, Morgan. 2019. "The Standard Errors of Persistence." June 3. Available at SSRN. DOI: `10.2139/ssrn.3398303`

Orlow, Dietrich. 1991. *Weimar Prussia, 1925-1933: The Illusion of Strength.* Pittsburgh: University of Pittsburgh Press.

Schattkowsky, Ralph. 1994. "Separatism in the Eastern provinces of the German Reich at the end of the First World War." *Journal of Contemporary History* 29: 305-324.

Schulz, Gerhard. 1963. *Zwischen Demokratie und Diktatur: Verfassungspolitik und Reichsreform in der Weimarer Republik. Band 1. Die Periode der Konsolidierung und der Revision des Bismarckschen Reichsaufbaus 1919-1930.* Berlin: de Gruyter.