Causal Effects, Migration, and Legacy Studies ▦ ⬡

Moritz Marbach   University College London

Abstract: *Political scientists have long been interested in the persistent effects of history on contemporary behavior and attitudes. To estimate legacy effects, studies often compare people living in places that were historically exposed to some event and those that were not. Using principal stratification, we provide a formal framework to analyze how migration limits our ability to learn about the persistent effects of history from observed differences between historically exposed and unexposed places. We state the necessary assumptions about movement behavior to causally identify legacy effects. We highlight that these assumptions are strong; therefore, we recommend that legacy studies circumvent bias by collecting data on people's place of residence at the exposure time. Reexamining a study on the persistent effects of U.S. civil rights protests, we show that observed attitudinal differences between residents and nonresidents of historic protest sites are more likely due to migration rather than attitudinal change.*

Verification Materials: The data and materials required to verify the computational reproducibility of the results, procedures and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/NGMCDS.

Political scientists have long been interested in the persistent effect of local historical events, policies, or institutional changes on contemporary political behavior and attitudes of individuals. At the heart of these studies is often a comparison of the attitudes and behavior of residents in historically exposed and unexposed places. For example, a recent study by Homola, Pereira, and Tavits (2020) shows that Germans living closer to former Nazi concentration camps today tend to be more xenophobic, intolerant, and more likely to vote for a far-right party. Another example is Rozenas and Zhukov (2019), who link Stalin's terror by hunger to Ukrainian citizens' loyalty and opposition toward Moscow across eight decades. A third example is Mazumder (2018) demonstrating that respondents living in counties where civil rights protests took place in the 1960s are less likely to harbor racial resentment against Blacks.[1]

How much can be learned about the causal effect of historical events on contemporary attitudes and behavior from observed differences between residents in historically exposed and unexposed places? To the extent that people can move after the historical event of interest, any comparison between contemporary outcomes of people in places historically exposed with people in places historically unexposed is not a comparison of exposed and unexposed people alone. For example, people living in historically exposed places today include people that lived in these places at the time of exposure but they also include people that lived elsewhere and may have not been subject to the exposure. Therefore, observed differences will not necessarily equal the causal effect of the exposure on individuals even if the exposure was completely randomly assigned across places. Authors of legacy studies recognize that migration possess a challenge to the interpretation of their results.

Moritz Marbach, Associate Professor, Department of Political Science, University College London, 29-31 Tavistock Square, London, WC1H 9QU, United Kingdom (m.marbach@ucl.ac.uk).
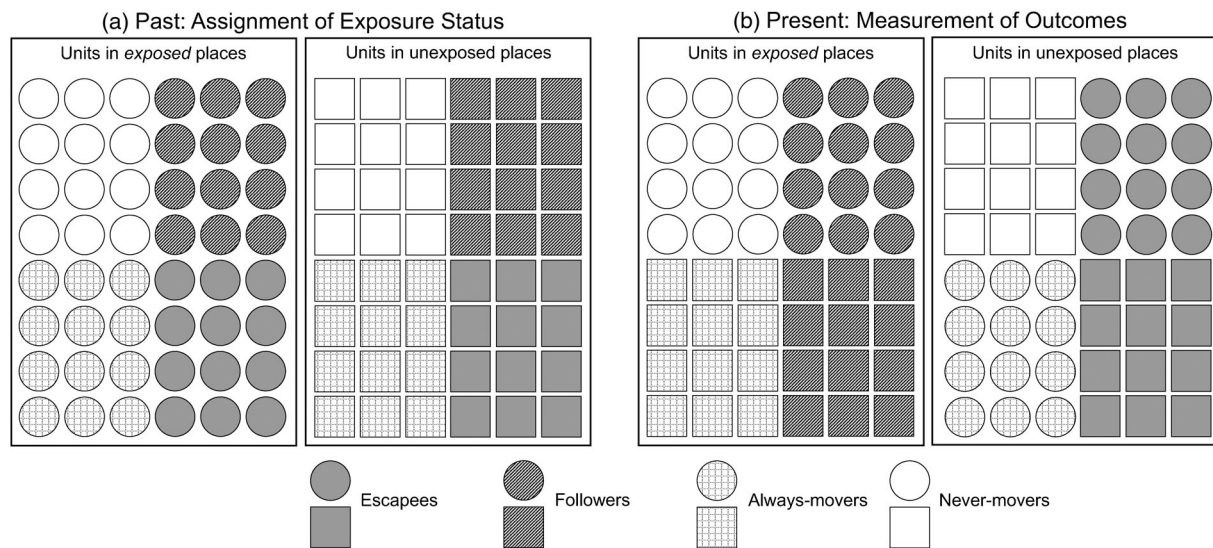
[1] For a recent review on legacy studies, see Cirone and Pepinsky (2021).

FIGURE 1 Example of Post-Treatment Sorting

Notes: At the time of exposure assignment, the distribution of four latent mover types is balanced between exposed and unexposed places (Panel A) but at the time of outcome measurement (after sorting), the distribution is unbalanced (Panel B). The unbalanced-type distribution is the reason for why the comparison of individuals in exposed and unexposed places differs from the comparison of exposed and unexposed individuals. Whereas the latter is the causal effect of the exposure, the former is distorted by post-treatment sorting bias. As the exposed and unexposed always-movers swap places, their presence dilutes the causal effect whereas the concentration of escapees and followers in exposed or unexposed places distorts the causal effect if these types are ex ante heterogeneous in their outcomes.

However, the problem is that we lack a formal framework to analyze how and when migration matters for our ability to identify the causal effect of historic events on contemporary outcomes.

In this article, we address this gap in the literature by providing such a framework. We focus on a setting in which the effect of a one-time shock on an individual's own outcomes is of interest.[2] Similar to the four complier types in instrumental variable estimation (Angrist, Imbens, and Rubin 1996), we define four latent mover types (never-movers, always-movers, escapees, and followers) and show how the presence of these mover types in the population leads to, what we refer to as, post-treatment sorting bias. We consider a setup in which an exposure status is randomly assigned across places and the distribution of latent mover types is balanced (see Figure 1). Once the exposure is realized, some types sort from exposed to unexposed places (and vice versa). Sorting, that is movement between exposed and unexposed places, leads to a distribution of types that is unbalanced

at the time outcomes are measured, which is the source of post-treatment sorting bias.

We demonstrate that even if sorting is not caused by the exposure (there are only never- and always-movers), the causal effect cannot be recovered without bias. The reason is that some individuals sort regardless of their exposure status (the always-movers) thus diluting the average treatment effect (ATE). In the example displayed in Figure 1, exposed always-movers swapped places with unexposed always-movers mixing exposed and unexposed individuals. We demonstrate that under general conditions (all four types are present), post-treatment sorting bias can be of any magnitude and go in any direction. The intuition is that escapees fleeing from the exposure are concentrating in the unexposed places whereas followers congregate in exposed places (see again Figure 1). If the potential outcomes from these two types differ ex ante, differences between exposed and unexposed places will be partially due to this heterogeneity.

The solution to post-treatment sorting bias is simple from a theoretical perspective: Studies should compare historically exposed and unexposed individuals instead of comparing individuals living in exposed and unexposed places. Although theoretically straightforward, the practical implementation requires that studies find data—or collect their own data—that include

[2]In the discussion section, we consider the role of sorting in settings in which the exposure is permanent rather than a one-time shock, exposure-outcome effects span generations and outcomes are not measured on the individual level but on the community (aggregate) level.

information on where someone lived at the time of the exposure. An example for this approach is a study by Barceló (2021), who examines the effect of war exposure on civic engagement. Rather than correlating civic engagement with the historic conflict intensity in respondents' contemporary place of residence, the author obtained data on respondents' place of residence approximately at the time of the war to construct the conflict-exposure variable. If authors are unable to obtain the required data, our article might help authors to discuss the magnitude and the direction of the sorting bias one may expect.

Throughout this article, we reanalyze a study on the persistent effect of the U.S. civil rights protests as a running example (Mazumder 2018). In this important study, Mazumder argues that attitudes of Whites toward Blacks shifted in counties that were scenes of civil rights protests in the 1960s. The empirical analysis relies on contemporary survey responses among White respondents in the cross-sections of 2006–11 from the Cooperative Congressional Election Study (CCES). More specifically, the analysis uses survey responses and respondents' current place of residence to construct county-level averages measuring political attitudes and regresses those on a binary indicator if any civil rights protests occurred in the same county between 1960 and 1965. Relying on a selection-on-observable assumption, the estimates suggest that survey respondents living in U.S. counties where civil rights protests occurred are more likely to support affirmative action, harbor less political resentment toward Blacks, and are more likely to identify as Democrats.

We use this study as a running example because the publicly available survey data provide additional information on respondents' place of residence when they were 10 and 17 years old. These additional data allow us to directly observe whether a subset of survey respondents were exposed to the protests in the 1960s. Using the historic exposure status in the main regressions eliminates most sorting bias by definition. Leveraging these new data, we find little evidence for a legacy effect of these protests on political resentment toward Blacks, support for affirmative action, or Democratic party identification. The absence of treatment effects suggests that the observed differences between protest and non-protest counties are due to post-treatment sorting bias. Consistent with this statement and the White-flight hypothesis (Duncan and Duncan 1957; Taeuber and Taeuber 1965), we find evidence that protests increased the likelihood of Whites leaving protest counties in the 1960s.

Most legacy studies including this one grapple with sorting. Typically, studies present evidence that

demographic measures such as net-migration or total population change are not differential between historically exposed and unexposed places and therefore, so the argument goes, cannot confound the treatment–outcome effect. Using measures of migration, studies also use mediation analysis to demonstrate that there is a direct effect of the exposure on the outcome. Using the framework, we discuss if these strategies help to isolate the causal effect from the post-treatment sorting bias. We also discuss the role of sorting in other settings, including settings in which the exposure is permanent rather than a one-time shock, exposure–outcome effects span generations, and outcomes are measured on the cluster rather than the individual level. The discussion highlights that sorting is not less important in these settings.

The main contribution of this article is that it conceptualizes and characterizes the consequences of unobserved post-treatment sorting in legacy studies analyzing political behavior and attitudes. We highlight that post-treatment sorting bias can be thought of as an instance of measurement-error bias. Relying on the principal strata framework (Frangakis and Rubin 2004), we discuss the difficulties of removing sorting bias based on behavioral assumptions about mover types. We detail a series of assumptions assuring that sorting bias attenuates the treatment effect toward zero and therefore mimics the bias due to nondifferential measurement error in the exposure variable (Aigner 1973; Lewbel 2007). We highlight that the necessary assumptions are stringent and often unrealistic. We, therefore, side with the increasing scepticism toward the convenient assumption of nondifferential measurement in realistic settings (e.g., Imai and Yamamoto 2010). This study also contributes to a growing methodological literature on how to isolate persistent effects from observational data (e.g., Kelly 2019); see also Voth (2021) for a review. Empirically, we contribute to the growing body of literature that highlights how internal migration shapes (and reshapes) political geography in the United States (e.g., Brown and Enos 2021; Cantoni and Pons 2022).

Unobserved Mover Types

We consider a population of N individuals. Let C_i denote an indicator to which cluster an individual i belongs and let Z_i denote the binary (treatment) exposure.[3] Throughout this article, we assume that the treatment is assigned at the cluster level (spatial treatment), such that the individuals in a cluster c are either all exposed or not exposed,

---

[3]The binary exposure setting is a useful starting point as it lends itself to the simple typology of mover types that we introduce here.

TABLE 1 Defining Four Latent Mover Types

Types	Exposure and Potential Movement
Always-mover	$M(0) = 1 \quad M(1) = 1$
Never-mover	$M(0) = 0 \quad M(1) = 0$
Escapee	$M(0) = 0 \quad M(1) = 1$
Follower	$M(0) = 1 \quad M(1) = 0$

Notes: The four latent mover types are defined based on their movement choices when exposed ($M_i(1)$) and when not exposed ($M_i(0)$). The movement choices are to either stay ($M(\cdot) = 0$) or move away ($M(\cdot) = 1$).

that is, $Z_c \in 0, 1$. If individual i lived in a cluster that was exposed (treatment cluster), the exposure of this individual takes the value 1. If the individual lived in a cluster that was not exposed (control cluster), the exposure of this individual takes the value 0. In our running example, a respondent is exposed if they lived in county with civil rights protests at the time when these protests happened.

We focus on a situation in which the exposure is a shock (e.g., protest events) rather than institutional change (e.g., a new local law). The key difference between these two types of exposures is that a shock is a one-time event whereas an institutional change persists over time. For the latter, the notion of exposed and unexposed is more complicated as individuals differ (possibly endogenously) in how long they are affected by the exposure. We return to this point in the discussion section below.

After the exposure is realized but before outcomes are measured, individuals may move from treatment to control clusters (and vice versa).[4] Because individuals' choice to move may depend on exposure, we define two *potential movement* choices for each individual: one for when the individual is exposed, $M_i(1)$, and one when it is unexposed, $M_i(0)$. Depending on an individual's exposure status, only one of the two movement choices is realized and, at least in principle, could be observed. The realized *movement*, M_i, measures if an individual decided to move ($M_i = 1$) or not ($M_i = 0$). In our setup, it is sufficient to consider moves from a treatment to a control cluster or from a control cluster to a treatment cluster (but ignore moves across treatment clusters or across control clusters).

Similarly to the compliance types defined by Angrist, Imbens, and Rubin (1996), we define four different *mover types* depending on individual potential movement choices (see Table 1). Individuals that always move

regardless of exposure status are referred to as *always-movers* whereas individuals that never move are characterized as *never-movers*. When an individual only moves when exposed, we refer to such an individual as *escapee*. Similarly, when an individual only moves when not exposed, we refer to such an individual as *follower*. For the last two types, the exposure prompts an individual to move: Escapees flee from treatment clusters whereas followers move toward treatment clusters. Let S_i be a variable encoding an individual's moving type: never-mover (N), always-mover (A), escapee (E), or follower (F).

Exposure and individuals' mover type jointly determine an individual's residency at the time when outcomes are measured. We define an individual's *residency* R_i, by measuring if an individual resides in a cluster (historically) exposed to the treatment ($R_i = 1$) or not ($R_i = 0$) at the time when outcomes are measured. More specifically, we can write R_i as a function of potential movement and exposure, that is, $R_i(Z_i) = Z_i \cdot (1 - M_i(Z_i)) + (1 - Z_i) \cdot M_i(Z_i)$, or in terms of realized movement and exposure, that is, $R_i = Z_i \cdot (1 - M_i) + (1 - Z_i) \cdot M_i$. In the civil rights protest study, R_i measures if respondents currently live in a county that was the site of protests in the 1960s (or not).

Let $Y_i(r, z)$ be the potential outcome for an individual i with exposure z and residency r. We define four instead of two potential outcomes to accommodate situations in which moving affects the outcome. Let Y_i be the realized outcome. For each individual, only one potential outcome is realized. Without loss of generality, we leave the time difference between exposure and outcome measurement implicit.[5]

We assume that one seeks to identify the ATE of Z, which is defined as $\delta = \mathbb{E}[Y(R(1), 1) - Y(R(0), 0)]$. We say that δ is a legacy effect as it is measured some time after the exposure occurred. Throughout this article, we abstract from any confounding and assume that the exposure is randomly assigned:

Assumption 1. *Exposure randomization*
$Y(0, 0), Y(1, 1), Y(1, 0), Y(0, 1) \perp\!\!\!\perp Z.$

In our running example, the author does not assume that protests are randomly assigned but rather that after conditioning on a series of observables, it is as if random

---

[4]We abstract from the possibility that individuals may die or move to a cluster that is neither treatment nor control (i.e., that there is attrition). In the civil rights protest study, that means, for example, that there is no emigration to other countries.

[5]The definition of the potential outcomes implies that there is no interference between individuals. In the setting considered here, the stable unit treatment assumption rules out, for example, that the potential outcomes of an unexposed never-mover j are dependent on the arrival of an exposed escapee i in the same cluster. In the running example, this assumption could be violated if, for example, White escapees affect the racial resentment of Whites at their destination.

if someone got exposed (or not) to these protests. To facilitate the intuition, we discuss sorting bias in the context of a completely randomized exposure rather than a conditional independent exposure.

The focus of this article is on the identifiability of the individual-level ATE. Not all legacy studies target this estimand. In cluster-level legacy studies, the target estimand is the difference between a cluster's potential outcome when exposed and a cluster's potential outcome when not exposed. We discuss the role of sorting in these cluster-level legacy studies in the discussion section.

Sorting Bias in Regressions

We assume that individuals sort if there are at least some individuals for which $R_i \neq Z_i$, that is, if at least some individuals move from any treatment cluster to any control cluster (or vice versa) *after* the exposure was realized. When individuals move from a treatment cluster to a control cluster (and vice versa) *before* the exposure is realized, we say that individuals select rather than sort. The focus of this study is on post-treatment sorting rather than pre-treatment self-selection.

With (cluster-level) randomization and information about individuals' exposure status at the time of exposure, the ATE is identified. For example, the coefficient δ estimates the ATE when regressing the observed outcomes (Y_i^{obs}) on the exposure indicator (Z_i):

$$Y_i = \alpha + \delta Z_i + \epsilon_i. \qquad (1)$$

However, as explained above, we assume that information about Z_i is not available and therefore, we compare treatment and control clusters instead (as defined by R_i):

$$Y_i = \alpha' + \delta' R_i + \epsilon_i'. \qquad (2)$$

In this regression, the coefficient δ' does not generally estimate the ATE. The estimates from such a regression are presented in the study on the persistent effect of the civil rights protests. The author regresses measures of political resentment on an indicator of whether a respondent lived at the time of a survey in a county with civil rights protests in the 1960s.

Post-treatment sorting bias is the difference $\delta' - \delta$. Conceptually, sorting bias originates in the imperfect observability of treatment exposure. Because an individual's residency (R_i) is a function of an individual's exposure (Z_i), we can think of R_i as a contaminated version of Z_i and rewrite the regression Equation (2) as a function of

TABLE 2 Latent Stratification of the Population, by Exposure and Realized Residency

		Z	
		0	**1**
R	0	Escapee & never-mover	Escapee & always-mover
	1	Follower & always-mover	Follower & never-mover

Notes: The columns represent the exposure (Z) and rows realized residency (R). Each cell contains two latent movers types.

realized movement and exposure:

$$Y_i = \alpha' + \delta'(Z_i(1 - M_i + (1 - Z_i)M_i) + \epsilon_i' \qquad (3)$$

$$= \alpha' + \delta' Z_i + \delta' M_i - 2\delta' Z_i M_i + \epsilon_i'. \qquad (4)$$

This regression equation highlights that regressing the observed outcome on the residency indicator is identical to regressing the observed outcome on (i) the treatment exposure indicator, (ii) the realized movement indicator, and (iii) the (scaled) interaction between the two while simultaneously constraining the coefficients for these three terms to the same value. Although this equation does not reveal much about the direction or magnitude of the bias, it highlights that δ' is unlikely to equal the ATE (here δ) in general.

Anatomy of Sorting Bias

We next decompose the δ' into a series of contrasts for the different mover types defined above. Different from the regressions in the previous section, we make no constant-effect assumption and use the mover typology introduced earlier. To facilitate the intuition of this decomposition, consider Table 2, which stratifies the population by realized residency (R) and exposure (Z). Each cell consists of individuals from two different types. Based on Z_i and R_i alone, these two different types per cell cannot be distinguished. For example, individuals in the upper left cell could be escapees (individuals that only move when exposed) or never-movers (individuals that never move no matter the exposure). In the upper right cell, individuals could be escapees (they left for a control cluster after getting exposed) or always-movers (individuals that always move no matter the exposure). The same logic applies to the cells in the lower row.

Maintaining the randomized-exposure assumption (A1), we can decompose the observed difference between treatment and control clusters, δ' (the difference between outcomes from individuals in the lower row and the outcomes from individuals in the upper row) as follows:

$$\delta' = \mathbb{E}[Y|R=1] - \mathbb{E}[Y|R=0] \tag{5}$$

$$= \big(p(S=N|R=1)\mathbb{E}[Y(1,1)|R=1, S=N]$$
$$\quad - p(S=N|R=0)\mathbb{E}[Y(0,0)|R=0, S=N] \big)$$
$$+ \big(p(S=A|R=1)\mathbb{E}[Y(1,0)|R=1, S=A]$$
$$\quad - p(S=A|R=0)\mathbb{E}[Y(0,1)|R=0, S=A] \big)$$
$$+ \big(p(S=F|R=1)\mathbb{E}[Y|R=1, S=F]$$
$$\quad - p(S=E|R=0)\mathbb{E}[Y|R=0, S=E] \big) \big). \tag{6}$$

This decomposition shows that the observed difference is the sum of three contrasts. The first contrast is the difference between exposed individuals and unexposed individuals among never-movers. In terms of Table 2, we compare individuals on the main diagonal. If the population would only include never-movers (i.e., if individuals were not sorting), we could attribute the difference between treatment and control clusters to the exposure. Yet, the presence of the other mover types complicates matters.

The second contrast is the difference between unexposed individuals and exposed individuals among the always-movers. Although this contrast also compares exposed and unexposed individuals, the direction of the comparison is flipped. Assuming that treatment effects are homogeneous, this second contrast is the treatment effect but with the opposite sign.

The third contrast is not a causal contrast but a comparison of observed outcomes among the escapees (of which some were exposed whereas others were not) and followers (of which some are exposed and some are not). This last difference characterizes the heterogeneity in outcomes between individuals that move *because* of the exposure.
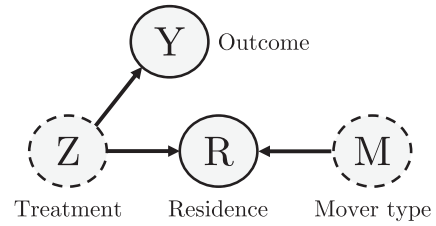
Random Sorting

Next, we outline three restrictive assumptions to ensure that sorting bias has the same consequences as nondifferential measurement-error bias, which attenuates the ATE toward zero (Lewbel 2007).

We show that sorting dilutes the average treatment if sorting (i) is not sparked by the exposure (A2, no escapees and followers), (ii) has no causal effect on the outcome (A3, incidental sorting), and (iii) types are homogeneous in their potential outcomes (A4, non-differential types). Under these three assumptions, the ATE is attenuated toward zero (diluted) as some individuals sort regardless of the exposure (the always-mover).

Formally, we write these three assumptions as follows:

FIGURE 2   Directed Acyclic Graph with Random Sorting Assumptions

Notes: Observed random variables are solid whereas unobserved random variables are dashed.

Assumption 2. *No escapees and followers* $M(1) = M(0)$.

Assumption 3. *Incidental sorting* $Y(0,1) = Y(1,1)$, $Y(0,0) = Y(1,0)$.

Assumption 4. *Partial non-differential mover types* $E[Y(1)|M=1] = E[Y(1)|M=0]$, $E[Y(0)|M=1] = E[Y(0)|M=0]$.

The corresponding causal graph appears in Figure 2.

The first assumption states that the exposure does not spark sorting. This is equivalent to say that there are no escapees or followers but only never- and always-movers. One implication of this assumption is that the realized movement indicator is sufficient to characterize individuals' types, that is, $M(0) = M(1) = M$. In other words, if $M_i = 1$, this individual is an always-mover, whereas the individual is a never-mover if $M_i = 0$.

The incidental sorting assumption (A3) rules out any effect of moving on individuals' outcomes if these individuals have the same exposure status. An implication of this assumption is that we can focus on two instead of four potential outcomes: One for the individual when exposed, $Y_i(1)$, and one for the individual when it was not exposed, $Y_i(0)$. Let Y_i be the realized outcome, which is either $Y_i(1)$ when the individual was exposed and $Y_i(0)$ otherwise.

In combination with the previous two assumptions, the non-differential mover-type assumption rules out any heterogeneity between the average potential outcomes of never- and always-movers. We assume that their average potential outcomes under treatment (and under control) are identical. A stronger but unnecessary version of this assumption is to assume that all types are non-differential $(Y(1), Y(0) \perp\!\!\!\perp S$ or $Y(1), Y(0) \perp\!\!\!\perp M(1), M(0))$.

Combining all three assumptions (A2–A4) and maintaining the randomization assumption (A1), we have:

$$\delta' = \mathbb{E}[Y(1) - Y(0)] \underbrace{(p(Z = 1|R = 1) - p(Z = 1|R = 0))}_{\text{attenuation term}}$$

$$= \delta\pi. \tag{7}$$

The proof appears in Appendix C.1 in the Supporting Information (pp. 11–12). Under the assumptions A1–A4, δ' is the product of δ and a constant π, which is bounded between -1 and 1.

In most instances, it is reasonable to assume that it is more likely to observe exposed rather than unexposed individuals in treatment clusters ($p(Z = 1|R = 1) > 0.5$) and that it is more likely to observe unexposed rather than exposed individuals in control clusters ($p(Z = 1|R = 0) < 0.5$). If so, π is bounded between 0 and 1.

The result highlights that even if the exposure does not spark movement, sorting is incidental, and mover types are non-differential in their potential outcomes, there is still a mixing of exposed and unexposed individuals in treatment and control clusters. This mixing dilutes the ATE. In the most extreme scenario, when there is complete mixing (it is equally likely to observe exposed and unexposed individuals in treatment and control clusters), π sets δ to 0. In more moderate scenarios, the ATE is attenuated toward 0. If the joined distribution of $p(Z, R)$ (or $p(M, R)$) is known (e.g., from census tabulations), one could correct for this bias by dividing the observed difference between historically exposed and unexposed places (δ') by π.

The availability of census tabulations brings other opportunities. Suppose census data allow us to verify that $p(Z = 1) = p(R = 1)$, that is, sorting does not alter the relative population distribution between exposed and unexposed places. In that case, A1–A3 are sufficient to show that $\delta' = \mathbb{E}[Y(1) - Y(0)] - 2p(M = 1)\mathbb{E}[Y(1) - Y(0)|M = 1]$ (the proof appears in Appendix C.2, in the Supporting Information, pp. 12–13). This result shows that incidental sorting without escapees and followers and a stable relative population distribution leads to a bias that is additive and equals the product of the causal effect among the always-movers scaled by the size of the always-mover subpopulation.[6]

---

[6]We are grateful to one of the anonymous reviewers for pointing us to this result.

Non-Random Sorting

Although all three assumptions (A2–A4) are highly restrictive, the assumption that there are no escapees and followers appears in particular implausible. In the context of the U.S. civil rights protest study, it amounts to assuming that no White residents moved from protest to non-protest counties because of the protest. Yet, a larger literature starting with Duncan and Duncan (1957) and (Taeuber and Taeuber 1965) suggests that the arrival of Blacks in historically White neighborhoods sparked White out-migration. Because the protest increased the salience of the Black community in protest counties, one might suspect that the civil rights protest could have had a similar effect on Whites.

Relaxing the assumption of no escapees and followers (A2) and combining the remaining assumptions, we can write δ' as follows:

$$\delta' = \mathbb{E}[Y(1) - Y(0)] \underbrace{(p(Z = 1|R = 1) - p(Z = 1|R = 0))}_{\text{attenuation term}}$$

$$+ \underbrace{(p(S = F)\mathbb{E}[Y|S = F] - p(S = E)\mathbb{E}[Y|S = E])}_{\text{sorting heterogeneity}}. \tag{8}$$

This result demonstrates that the difference between treatment and control clusters (δ') is the sum of an attenuated estimate of the ATE (as discussed above) and an additive bias term that we refer to as sorting heterogeneity. The proof appears in Appendix C.3 in the Supporting Information (p. 13).

The sorting heterogeneity is the observed difference describing the heterogeneity between two different subpopulations rather than a causal effect. Note that the resulting contrast compares two weighted means (but the weights do not sum to 1). The weighting makes the substantive interpretation of this contrast difficult in practice. However, suppose that the outcome is (on average) identical across the two subpopulations (this amounts to assuming that all types are non-differential, see above). Despite the absence of any heterogeneity, differences in the size of the follower and escapee subpopulation produce an observed difference. In fact, if the share of followers is larger than the share of escapees, the bias due to the sorting heterogeneity is positive whereas it is negative if there are more escapees than followers.

Consistent with the White-flight hypothesis, we might assume that there are escapees but no followers, which means that the bias due to the sorting heterogeneity is negative. This assumption amounts to a monotonicity assumption, stipulating that the exposure either has no effect on an individual or a strictly positive effect. Although such a monotonicity assumption has been

used in other contexts—such as instrumental variable estimation (Angrist, Imbens, and Rubin 1996)—it seems plausible that there are at least some individuals living in non-protest counties that decided to move to protest counties because of the protests. In other words, it seems plausible that some Whites were attracted by the protests, as they signaled, for example, openness and tolerance.

The previous discussion assumes that the outcomes of escapees and followers are on average identical. What if they are not? In the context of the civil rights protest study, it seems plausible to assume that there is outcome heterogeneity among followers and escapees: Those that left counties because of the protests (escapees) might harbor more racial resentment against Blacks today as compared with those that moved to these counties. Assuming that the two subpopulations are of equal size such heterogeneity in outcomes leads to a negative additive bias. If the two subpopulations differ in size and there are more escapees than followers, the magnitude of the bias increases.

With non-random sorting, we cannot comfort ourselves that we merely underestimate the treatment effect as in a setting with random sorting. The key concern for applied research is that the true exposure effect might be zero and all differences between treatment and control clusters are pure sorting heterogeneity. Because the bias due to the sorting heterogeneity is additive, observed differences between treatment and control clusters provide little information even about the sign of the exposure effect.

Removing Sorting Bias by Design

The previous discussion highlights that sorting bias is a threat to causal inference. The solution is, from a theoretical perspective, simple: Using data on people's exposure status instead of relying on current residency as a proxy avoids sorting bias. For legacy studies based on survey data, this means either constructing an exposure variable based on information where respondents lived at the time of the treatment or directly eliciting respondents' exposure status. The study by Barceló (2021) mentioned in the introduction is an example of the first approach, whereas the study by Lupu and Peisakhin (2017) is an example of the second approach.[7]

We illustrate the first approach by reexamining the civil rights protest study of Mazumder (2018).[8] Our extended analysis uses data from the CCES 2010–14 Panel Study (Ansolabehere and Schaffner 2015). This survey re-interviewed the 8014 White respondents from the CCES 2010 study in 2012 and 2014. Importantly, in 2012, respondents were asked in which town or city they lived on their 17th birthday, and in 2014 in which town or city they lived on their 10th birthday. Responses from these two questions allow reconstructing for some birth cohorts in which county they lived between 1960 and 1965, and more specifically if they lived in a county where protests took place. The analysis sample includes White respondents and excludes everyone that did not live in the contiguous United States during the times of the protests ($N = 3,244$). A description of the survey items and how we identified respondents' county of residence can be found in Appendix A in the Supporting Information (pp. 2–3).

We follow the original study for the baseline identification strategy and assume that the treatment exposure (protests) is as if randomly assigned on the county level, conditional on a series of covariates. The relevant covariates include: the average democratic vote share between 1932 and 1960, the logarithm of the county population as measured in 1960, median income in 1960, the percentage of Blacks among the total population in 1960, as well as the percentage in urban areas in 1960. For a few respondents, no information on their place of residency in the 1960s is available, which reduces the number of observations in the regressions below.

The main outcome is a continuous, positive summary score from a principal component analysis (PCA) of the average response across three panel waves of the survey items used in the study of Mazumder.[9] Higher values on this variable indicate more resentment toward Blacks (the item correlations are reported in Appendix Table SM.2, in the Supporting Information, p. 5). We prefer the PCA score, as it reduces measurement error and makes the presentation of the results more concise. However, in the Appendix Tables SM.3–SM.6 in the Supporting Information (pp. 6–9), we also report all regressions on each survey item separately.

---

[7]A concern with the suggested solution is that survey responses come with measurement error, which also biases estimates. Therefore, it is important to construct survey instruments that minimize measurement error. Which of the two strategies performs better in practice in that regard remains an empirical question to be studied. In order to avoid measurement error altogether, it would be nec-

essary to combine the survey data collection with, perhaps, linked administrative data that include validated information on respondents' past addresses.

[8]See Biggs, Barrie, and Andrews (2020) for another discussion on this study.

[9]To allow for a concise discussion of the results below, we construct the PCA score to be strictly positive by shifting all values with some constant.

TABLE 3 Proportion of White CCES Respondents, by Place of Residency

		Lived in Protest County	
		No	**Yes**
Living in protest county	**No**	0.38	0.24
	Yes	0.11	0.27

Notes: Proportion of White respondents by residency at the time of the survey (2012/2014) and place of residence at the time of the protests (1969-65). $N = 3{,}244$.

Table 3 shows the number of respondents living in protest counties during 1960s and where they live today. We see that about 24% the respondents living in a protest county at the time of the protest moved to a non-protest county at the time of the survey. On the other hand, about 11% of the respondents did not live in a protest county but live in such a county now. Overall, we see that about a third of the respondents moved from protest to non-protest counties (and vice versa). Although we cannot distinguish if the movers are always-movers, escapees, or followers, it is obvious that there is ample room for post-treatment sorting bias.

Table 4 (column 1) replicates the baseline finding using the full sample.[10] Respondents living in counties that were the scene of civil rights protests in 1960–65 are much less resentful toward Blacks than those living in counties where no such protests occurred. The estimated effect suggests that respondents living in a protest county have a 0.14 lower PCA score than those living elsewhere. This effect is about 7.8% of a standard deviation in the PCA score.

The second column of the same table replicates the baseline results for the sample of respondents for which information about their place of residence in 1960–65 is available. The estimated effect in this subsample is larger suggesting that there is effect heterogeneity by birth cohort. In the third column, we report the estimates from a regression where instead of including state fixed-effect for respondents' current county of residence and covariates about the attributes of their current county of residence, we include a fixed-effect for the state of the county respondents lived in and covariates on the attributes of their county of residence in the 1960s. This adjustment is necessary to remove some of the post-treatment bias that

comes from conditioning on post-treatment covariates. The estimated effect is a bit larger again.

The fourth column includes an indicator if the respondent lived in a protest county during the time of the protests instead of an indicator if the respondent now lives in a protest county. In this specification, the estimated effect is zero and the standard error is about 0.1. This means that respondents who experienced protests during 1960–65 have about the same level of racial resentment toward Blacks today as respondents that did not experience the protests in 1960–65. The estimate in the fourth column suggests that much of the differences between respondents living and not living in protest counties now are due to post-treatment sorting.[11]

The estimate in the fourth column is also consistent with another interpretation: The indicator for protest exposure in the 1960s was measured with error and, therefore, the estimate of the effect is attenuated to zero. There are two reasons that speak against this interpretation. First, the indicator is not based on respondents' self-reported exposure to the protests but based on their reported place of residence during adolescence. Although retrospective survey questions always come with the risk of recall error, reporting one's place of residence during adolescence seems less demanding than recalling protest exposure. Second, if the indicator for protest exposure in the 1960s was noise alone, we would expect to never find any effects. However, the next result shows that this is not the case.

The fifth column shows that the protests were a strong push factor for Whites. In the last column of Table 4, we regress an indicator if a respondent lived in the same type of county between 1960 and 1965 and in 2010 (either always in a treatment county or always in a control county) on the indicator for protests in 1960–65. The estimated coefficient suggests that the chances to remain in the same county-type is about 23% lower for those that lived in protest counties in 1960–65, as compared with those in the non-protest counties. This suggests that the protests induced White flight. However, it is important to keep in mind that this result only means that there are relatively more escapees (White fleeing from protest counties) than followers (White attracted to move to protest counties).

The results demonstrate that the sorting heterogeneity (the second under-braced term in Equation (8)) is

---

[10]In the Supporting Information, we present estimates based on county-level aggregates following the original study (see Appendix Table SM.8 in the Supporting Information, p. 10). We also demonstrate that the estimates are similar when using the provided survey weights (see Table SM.7 in the Supporting Information, p. 9).

[11]The absence of an ATE may mask treatment effect heterogeneity by latent mover type. In principle, it is possible that there is a positive treatment effect among some types (e.g., the never-movers) and a negative treatment effect among any of the other types (e.g., the escapees). We leave it to future research to devise a means to estimate treatment effect heterogeneity by latent mover type.

TABLE 4 OLS Regression Estimates of the Effect of Living in a Protest County

	Resentment PCA Score				Stays?
	Model 1	Model 2	Model 3	Model 4	Model 5
Lives in protest county?	$-0.14^{\dagger}$	-0.28^{**}	-0.30^{**}		
	(0.08)	(0.10)	(0.07)		
Lived in protest county 1960-65?				-0.00	-0.23^{**}
				(0.10)	(0.03)
N	7,929	3,212	3,216	3,216	3,216
Residence county					
State FE	Yes	Yes	No	No	No
Covariates	Yes	Yes	No	No	No
Residence county 1960–65					
State FE	No	No	Yes	Yes	Yes
Covariates	No	No	Yes	Yes	Yes

Notes: The dependent variable is the racial resentment summary score from a principal component of four survey items measured in 2010-2014 (col. 1-4) and an indicator whether a respondent lived in a protest (non-protest) county at the time of the protests and at the time of the survey (col. 5). Higher values indicate more racial resentment. The sample includes all CCES respondents (col. 1) and all CCES respondents that were ten or 17 years old in 1960-65 when the protests took place (col. 2-5). Cluster-robust standard errors at the county level in parentheses.
Significance level: $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

negative. This is consistent with the hypothesis articulated in the previous section that the respondents who left protest counties because of the protests (the escapees) have on average higher resentment toward Blacks as compared with those that moved to these counties (the followers). The fact that there are more escapees than followers amplifies the difference and therefore increases the magnitude of the negative bias.

Existing Strategies to Address Sorting

Most legacy studies, including the civil rights protest study, grapple with sorting in one way or another. Depending on data availability, legacy studies use three different strategies to address sorting. The first strategy is to evaluate if migration flows are differential between historically exposed and unexposed places. In the absence of any differences, authors argue that sorting is not a concern. From the perspective of the proposed framework, we can think of this strategy as seeking to estimate the share of escapees and followers. Their presence is a necessary condition for a bias due to sorting heterogeneity. Estimating the share of escapees and followers is feasible if data on in-migration and out-migration are available. However, in practice, authors often only have access to data on net-migration (in-migration minus out-migration), which is only partially informative because a non-differential net-migration may mask an equally sized but sizeable share of escapees and followers. A case in point is the example in Figure 1, where net-migration is exactly zero, but sorting remains a concern. The civil rights protest study documents that the White net-migration rate based on U.S. Census records were slightly lower in protest counties relative to non-protest counties during the 1960s and the 1970s (the difference is only statistically significant in the 1970s). This is consistent with the finding from the analysis in the previous section that there are more escapees relative to followers. However, different from this analysis, the pattern documented in the civil rights protest study is limited to the two decades after the protests. Even if the net-migration rates were exactly equal during these two decades, there would still be ample room for sorting up until when outcomes are actually measured.

The second strategy used by legacy studies to address sorting is to conduct a mediation analysis using the data on migration rates. Prompted by the statistical significant differences in net-migration rates during the 1970s, the author of the civil rights protest study uses this strategy. The results of the causal mediation analysis (Imai et al. 2011; Imai, Keele, and Yamamoto 2010) indicate that the average direct effect (ADE) equals almost exactly the estimated treatment. The author then concludes: "As

a result, it does not seem that sorting alone can explain the entirety of the results." That conclusion, however, is premature. The estimated ADE describes the effect of the protests when the net-migration rate was held constant at the value that would have been realized in a world without protests. In other words, the ADE describes how a community-level variable mediates the effect of the exposure. Although the absence of the estimated mediation effect is an interesting result on its own, it does not help to isolate the average causal effect from the bias that comes with the individual sorting behavior. It is, of course, equally premature to conclude that mediation analysis is ineffective for all legacy studies. For cluster-level legacy studies, mediation analysis delivers the relevant causal quantity as we discuss in the next section.

A third and final strategy used by legacy studies to address sorting is to use rich census data to compare correlates of political behavior and attitudes of movers and nonmovers. An example of this approach is the study by Acharya, Blackwell, and Sen (2016), who show that contemporary racial resentment toward Blacks and opposition to affirmative action in Southern U.S. counties goes back to the prevalence of slavery 150 years ago. For data availability reasons, they study the demographics of movers and nonmovers in the 1930 U.S. Census, writing: "If sorting plays an important role in our results, we would expect to see differences between migrants to/from high-slave areas versus low-slave areas" (p. 630). In terms of Table 2, we can think of their comparison as contrasting the difference between movers (the off-diagonal cells) versus nonmovers (the diagonal cells) in each column. However, because each cell in the table is populated by two distinct unobserved types, their contrast does not isolate a (relative) difference between any of the four mover types. Although this does not invalidate their analysis, the absence of any difference does not imply that there is no post-treatment sorting, which originates in outcome differences between the four mover types. If it were credible to assume that one of the four mover types does not exist, one could use ideas from complier profiling in instrumental variable estimation to characterize the remaining three types (Marbach and Hangartner 2020). However, as discussed above, invoking a monotonicity assumption seems difficult to justify.

Extensions to Other Settings

There are a large variety of legacy studies. Surveying all legacy studies recently reviewed by Cirone and Pepinsky (2021), we find that the overwhelming majority of legacy studies examines exposures that vary across geography (about 86%). Among these studies, about a quarter focus on the effect of a one-time shock on individual-level outcomes—the setting we focus on in this article. In this discussion section, we discuss the role of sorting in other settings, including settings in which the exposure is permanent rather than a one-time shock, exposure-outcome effects spanning generations, and outcomes that are measured on the cluster, rather than the individual level. The discussion highlights that sorting is not less important in these settings.

Some legacy studies focus on exposure effects on descendants rather than on an individual's own outcomes. Although the framework presented in this article does not directly apply, the obstacles in identifying causal effects in such a setting are similar because a person's ancestors also have agency about their location choices. Empirically, the main difficulty is to track residence patterns for someone's multiple ancestors.[12] Theoretically, the difficulty is to define how the exposure status of a person's multiple ancestors maps into the exposure status of an individual.

In the one-time shock setting we focus on, individuals' assigned exposure is identical to individuals' exposure intake. Therefore, individuals are deprived of their ability to select their exposure after assignment by moving. For example, in the one-time-shock setting, followers moving to exposed places remain unexposed and escapees moving to unexposed places remain exposed. However, in some settings, individuals have the ability to select their exposure intake (or exposure length) by moving after assignment.[13] In settings in which the exposure is permanently changing institutions, policies, or socioeconomic features, individuals' agency over their own exposure intake via moving further complicates the identifiability of the exposure-outcome effect. From a practical perspective, this means that identifying the legacy effects of, for example, institutional change (a permanent exposure) may be more difficult as compared with legacy effects of an event (a one-time shock).

The setting we consider is one in which the causal estimand is the difference in the individual potential outcome when exposed and the potential outcome when

---

[12]Multigenerational surveys are rare but not unheard of in the literature on legacy studies. For example, Lupu and Peisakhin (2017) conduct a multigenerational survey to show that "descendants of individuals who suffered more intensely [from the deportation of Crimean Tatars in 1944] identify more strongly with their ethnic group, support more strongly the Crimean Tatar political leadership, hold more hostile attitudes toward Russia, and participate more in politics."

[13]In these settings, the exposure is a function of the latent mover type, that is, $Z(S)$.

not exposed. However, as already pointed out earlier, not all legacy studies target this quantity. What is the role of sorting in legacy studies focusing on the effects of an exposure on cluster-level outcomes rather than individual-level outcomes? In cluster-level legacy studies, the estimand is the difference between a cluster's potential outcome when exposed and a cluster's potential outcome when not exposed. Under random assignment of the exposure, these cluster-level differences are causally identified. Although cluster-level randomization of the exposure is sufficient to identify the exposure-outcome effect, the identification challenge in these studies is to discern if there is a causal effect net of the exposure effect on sorting.

We can cast the identification challenge in cluster-level legacy studies in terms of a causal mediation problem. Let $Y_j'(o, n, z)$ be the potential outcome for a cluster j with exposure z, out-migration level o, and in-migration level n. We also define potential out-migration and in-migration levels for cluster j given exposure z, that is, $O_j(z)$ and $N_j(z)$. Under random assignment, the average causal effect (total effect), that is, $\mathbb{E}[Y'(O(1), N(1), 1) - Y'(O(0), N(0), 0)]$, is identifiable. The challenge is to identify the average natural direct effect (NDE), that is, $\mathbb{E}[Y'(O(0), N(0), 1) - Y'(O(0), N(0), 0)]$, which describes the effect of the exposure on cluster-level outcomes if out- and in-migration were set to the level without the exposure. A growing literature discusses the causal identification assumption for NDEs that often involve difficult-to-justify cross-world comparisons (see, e.g., Imai et al. 2011; Imai, Keele, and Yamamoto 2010).

Conclusion

How much does history shape contemporary political behavior and attitudes? A number of studies suggest that the answer is "a lot." However, many of these studies compare individuals living in places that were historically exposed to some treatment with individuals living in unexposed places. As we demonstrate in this article, the observed differences can only be interpreted as a causal effect of the exposure on individuals if a number of strong identifying assumptions are invoked about how individuals migrate between historically exposed and unexposed places. If these identifying assumptions are not met, estimates will be contaminated by post-treatment sorting bias.

The fundamental problem of causal inference is that counterfactuals can never be observed. From this perspective, post-treatment sorting is a much smaller problem as the bias can be ruled out by design in principle with information about individuals' historic exposure status. This suggests that the best course of action for applied research is to find and collect better data. As demonstrated in this study, information about where survey respondents lived in the past is critical to obtain reliable estimates about how history shapes political behavior and attitudes.

The results in this article may help to guesstimate the magnitude and direction of the bias. One critical take-away point is that even in the best case scenario (random sorting), we should expect the treatment effect to be biased toward zero because the presence of always-movers dilutes the treatment effect. Moreover, the longer the period between treatment and outcome measurement, the higher the attenuation bias. As time passes, more and more individuals will move regardless of their exposure status. This means that the attenuation of the treatment effect should be expected to grow as time passes. Therefore, we may be in particular skeptical about studies demonstrating that the causal effects are dynamically increasing as time passes.

When using the results presented in this article to guesstimating the magnitude and direction of the bias, it is important to defend the underlying assumptions. That might not always be easy. For example, defending the incidental sorting assumption is difficult as the literature on the effects of moving on political behavior and attitudes in advanced democracies is rather thin and conflicting. Although some studies suggest that movers adopt the political ideology of their neighbors (e.g., Gallego et al. 2016), others find no effects on immigration attitudes among movers to urban areas (e.g., Maxwell 2019). Legacy studies would clearly benefit from more research on how domestic migration shapes, and is shaped by, political behavior and attitudes.

Another lesson from this article is that there are advantages of studying the effects of history with exposures that vary across groups rather than geography. Assuming that group membership is not subject to individual choice, the type of post-treatment bias outlined in this article is avoided. An example of this approach is the paper by Nunn and Wantchekon (2011) who study the effect of the slave trade on trust in Africa. Using contemporary survey data, their main analysis avoids the type of post-treatment sorting bias described in this article because the authors collected data on the intensity of the slave trade by ethnic group rather than (only) by geography.

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "The Political Legacy of American Slavery." *Journal of Politics* 78(3): 621–41.

Aigner, Dennis J. 1973. "Regression with a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics* 1(1): 49–59.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434): 444–55.

Ansolabehere, Stephen, and Brian Schaffner. 2015. "Cooperative Congressional Election Study: 2010-2014 Panel Study. Data Release No. 2." https://doi.org/10.7910/DVN/TOE8I1.

Barceló, Joan. 2021. "The Long-Term Effects of War Exposure on Civic Engagement." *Proceedings of the National Academy of Sciences* 118(6), e2015539118.

Biggs, Michael, Christopher Barrie, and Kenneth T. Andrews. 2020. "Did Local Civil Rights Protest Liberalize Whites' Racial Attitudes?" *Research & Politics* 7(3): 2053168020914757.

Brown, Jacob R., and Ryan D. Enos. 2021. "The Measurement of Partisan Sorting for 180 Million Voters." *Nature Human Behaviour* 5(8): 998–1008.

Cantoni, Enrico, and Vincent Pons. 2022. "Does Context Trump Individual Drivers of Voting Behavior? Evidence from U.S. Movers." *American Economic Review* 112(4): 1226–72.

Cirone, Alexandra, and Thomas B. Pepinsky. 2021. "Historical Persistence." *Annual Review of Political Science* 25(1): 241–259.

Duncan, Otis Dudley, and Beverly Duncan. 1957. *The Negro Population of Chicago: A Study of Residential Succession.* Chicago, IL: University of Chicago Press.

Frangakis, Constantine E., and Donald B. Rubin. 2004. "Principal Stratification in Causal Inference." *Biometrics* 58(1): 21–29.

Gallego, Aina, Franz Buscha, Patrick Sturgis, and Daniel Oberski. 2016. "Places and Preferences: A Longitudinal Analysis of Self-selection and Contextual Effects." *British Journal of Political Science* 46(3): 529–50.

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(2): 573–90.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105(4): 765–89.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25(1): 51–71.

Imai, Kosuke, and Teppei Yamamoto. 2010. "Causal Inference with Differential Measurement Error: Nonparametric Identification and Sensitivity Analysis." *American Journal of Political Science* 54(2): 543–60.

Kelly, Morgan. 2019. "The Standard Errors of Persistence." Working Paper WP19/13, UCD Centre for Economic Research.

Lewbel, Arthur. 2007. "Estimation of Average Treatment Effects with Misclassification." *Econometrica* 75(2): 537–51.

Lupu, Noam, and Leonid Peisakhin. 2017. "The Legacy of Political Violence across Generations." *American Journal of Political Science* 61(4): 836–51.

Marbach, Moritz, and Dominik Hangartner. 2020. "Profiling Compliers and Non-Compliers for Instrumental Variable Analysis." *Political Analysis* 28(3): 435–44.

Maxwell, Rahsaan. 2019. "Cosmopolitan Immigration Attitudes in Large European Cities: Contextual or Compositional Effects?" *American Political Science Review* 113(2): 456–74.

Mazumder, Soumyajit. 2018. "The Persistent Effect of US Civil Rights Protests on Political Attitudes." *American Journal of Political Science* 62(4): 922–35.

Nunn, Nathan, and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review* 101(7): 3221–52.

Rozenas, Arturas, and Yuri M. Zhukov. 2019. "Mass Repression and Political Loyalty: Evidence from Stalin's 'Terror by Hunger'." *American Political Science Review* 113(2): 569–83.

Taeuber, Karl E., and Alma F. Taeuber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change.* Chicago, IL: Aldine Publishing.

Voth, Hans-Joachim. 2021. "Persistence–Myth and Mystery." In *The Handbook of Historical Economics.* eds. Alberto Bisin and Giovanni Federico, San Diego: Academic Press, Elsevier, 243–67.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: Data Description
Appendix B: Further Empirical Results
Appendix C: Proofs