

Lines of Demarcation: Causation, Design-Based Inference, and Historical Research

Matthew A. Kocher and Nuno P. Monteiro

Qualitative historical knowledge is essential for validating natural experiments. Specifically, the validity of a natural experiment depends on the historical processes of treatment assignment and administration, including broader macro-historical dynamics. But if validating a natural experiment requires trust in the ability of qualitative evidence to establish the causal processes through which the data were generated, there is no good reason for natural experiments to be considered epistemically superior to historical research. To the contrary, the epistemic status of natural experiments is on a par with that of the historical research on which their validation depends. They are two modes of social-scientific explanation, each with its own pros and cons; neither is privileged. We illustrate this argument by re-examining an important recent contribution to the literature on violent conflict: Ferwerda and Miller's 2014 natural experiment estimating the causal effect of the German decision to devolve authority to the Vichy French government on violent resistance during World War II.

Political science is a theoretically and methodologically eclectic field of study that has traditionally drawn on influences from across the social sciences. It has also witnessed recurrent efforts to discipline

this diversity by specifying methodological hierarchies and drawing lines of demarcation between scientific and non-scientific research methods. The current trend toward “design-based inference” (DBI) is one such attempt, seen by its advocates as the leading edge of a revolution in the empirical analysis of politics.

DBI treats the randomized control trial (RCT) as the ideal research design to test “nomothetic” (law-like) generalizations about causal effects. For practical or ethical reasons, however, many important variables cannot be subjected to manipulation, limiting the applicability of experiments to many questions of great political interest. When this is the case, DBI extends the logic of experiments to observational research, judging its reliability based on how closely it approximates the experimental ideal. Specifically, DBI prescribes that researchers seek to identify “natural experiments”: instances in which a natural or social process has assigned units to distinct conditions in a way that is “as good as random.” When the as-if random assumption is satisfied, natural experiments permit the rigorous estimation of causal effects for a wider set of phenomena—and using a wider array of data sources—than would be possible using RCTs alone. Natural experiments thus represent an attractive *via media* between the narrow applicability of true experiments and the broad but messy abundance of observational data.

In the absence of actual or as-if randomization—when it is impossible to fully “distinguish the effects of causes from the effects of the conditions under which they operate”¹—proponents of DBI are skeptical about

A list of supplementary materials provided by the authors precedes the references section.

Matthew Adam Kocher is a lecturer in the Department of Political Science and the Jackson Institute at Yale University (matthew.kocher@yale.edu). Nuno P. Monteiro is Associate Professor of Political Science at Yale University (nuno.monteiro@yale.edu). The authors contributed equally to this article. For research assistance, they thank Bernard Bèzes, Benjamin Billingsley, Gabriel Botelho, Austin Carder, Fabiola Davila, Ajua Duker, Sarah Holder, Cassidy Lapp, Aube Rey Lescure, Sona Lim, Nils Metter, Usha Rungoo, Lyndon Sam, Angelina Xing, and, especially, Michael Repas. For comments and suggestions, they thank the editor of *Perspectives on Politics*, three anonymous reviewers, Ana Arjona, Peter Aronow, Kate Baldwin, David Collier, Allan Dافoe, Samuel DeCanio, Thad Dunning, Francesca Grandi, Timothy Guinnane, Greg Huber, Sigrun Kahl, Stathis Kalyvas, Audrey Latura, Adria Lawrence, Peter Liberman, Luis Schiumerini, Duncan Snidal, and participants in the Program on Order, Conflict, and Violence at Yale. A special note of gratitude is owed to John Mearsheimer for his unflinching support throughout the process of publishing this paper.

our ability to identify causal effects. At the bottom of the methodological hierarchy endorsed by DBI advocates lies “idiographic” historical research, which aims at tracing in detail the processes connecting causes with effects in particular cases. In contrast to research designs that aim at testing nomothetic propositions, idiographic research lacks an explicitly-defined population of cases that can serve as counterfactuals. Therefore, from the DBI perspective, idiographic research is unable to retrieve causal sequences, placing it outside the line of demarcation establishing the scope of scientific work.²

We argue that natural experiments depend on idiographic research. Specifically, a natural experiment can be considered valid only on the basis of qualitative evidence regarding the causal processes through which (i) units were assigned to treatment and control groups and (ii) the treatment was administered. Without such evidence, it is impossible to rule out alternative explanations for observed differences in the outcome. Statistical evidence, in the form of balance tests on observed covariates, is insufficient to validate natural experiments for the same reason that advocates of DBI deem matching and model-based statistical inference suspect: it is impossible to control for the theoretically indeterminate set of possible confounds.³ In this precise sense, testing nomothetic causal propositions using natural experiments depends logically upon idiographic causal inference about the particular contexts in which natural experiments are embedded. Therefore, the credibility of causal inferences based on natural experiments presupposes epistemic trust in the ability of historical research to establish causation—specifically, to establish that the relevant historical chains of events *caused* the data in such a way as to configure a natural experiment.

Once the crucial role of qualitative historical research in validating natural experiments is properly understood, however, there is no reason to privilege the epistemic status of the nomothetic propositions they support over that of the idiographic accounts necessary to support them. The trust we place in qualitative causal inferences should not depend on whether or not they are used to justify the assumptions of a natural experiment. Either idiographic causal inference *is* possible in general, in which case both historical research and natural experiments produce scientific knowledge; or idiographic causal inference is *not* possible, and neither historical research nor natural experiments are, in general, able to produce scientific knowledge. Either both historical research and natural experiments are inside the line of demarcation delimiting the scope of scientific discourse, or they are both outside. In our view, both natural experiments and conventional qualitative historical analysis have their strengths. Natural experiments allow for precise quantitative estimates of the local causal effects of specific variables, while historical research allows for a thicker qualitative

understanding of causal sequences and mechanisms. They are two different modes of producing scientific knowledge.

Beyond its skepticism of idiographic research, DBI shares with the previously dominant regression-centric approach to political methodology—epitomized by King, Keohane, and Verba’s (KKV) *Designing Social Inquiry*—a commitment to testing general causal propositions using quantitative techniques.⁴ Still, DBI departs from KKV’s focus on “external” validity and its emphasis on collecting as large a set of observations as possible. Following from the experimental focus on “internal” validity, DBI stresses knowledge of (or better yet, control over) the data generating process—knowledge that is particularly difficult to acquire in traditional macro-comparative, “large-N” research. Consequently, DBI tends to privilege sub-national, micro-comparative, or individual level research. While empirical reductionism is not a *necessary* feature of DBI, it is the dominant trend in contemporary political science.

By emphasizing the role of historical research in validating natural experiments, we also highlight the perils of reductionism in micro-level DBI. We argue that the historical account necessary to validate a natural experiment must encompass all the relevant levels of analysis, including broader macro-level dynamics shaping the data. This layering of levels of analysis is, in fact, one of the advantages of historical research; and it is key to validating a natural experiment.

While some proponents of natural experiments acknowledge the value of “shoe-leather” qualitative research,⁵ we have few examples in the literature that evaluate existing natural experiments against historical evidence, “perhaps precisely because of the extensive case knowledge and detailed information required to do so.”⁶ After presenting our general critique of DBI and laying out our abstract case for idiographic causal inference, we illustrate our argument empirically by re-examining a recently-published piece of scholarship that epitomizes the promise and peril of DBI research.

In “Political Devolution and Resistance to Foreign Rule: A Natural Experiment,” Jeremy Ferwerda and Nicholas L. Miller (henceforth, FM) exploit a purported natural experiment created when Germany conquered France in 1940 to test a general causal proposition about the effect of devolving political power to local elites on resistance to foreign occupation.⁷ For over half of World War II (WWII), Germany occupied and administered the northern and western portions of the country, but the south and southeast escaped occupation and were reconfigured as a diminished French state ruled from Vichy. The boundary dividing the two zones was called the Line of Demarcation (LoD), which FM say was placed as-if randomly at the local level. While the Germans were highly influential in both zones, FM claim that a greater degree of authority was “devolved” upon the French in Vichy than in the occupied zone, making

German rule in the occupied zone more direct than it was in Vichy. Political devolution, they argue, affected French motivations to resist. Because Vichy was ruled by a right-wing government, resistance there was largely conducted by the left, while in the occupied zone, Frenchmen of all political stripes were motivated to resist. Testing their argument in four French departments, FM find that close to the LoD more acts of resistance occurred in the occupied zone. Since they believe that distinct levels of devolution of power applied in the two zones independently of their “pre-treatment” characteristics, FM claim to have identified the causal effect of devolution of power on violent resistance to foreign occupation.⁸

Given the modest external validity of a test conducted on less than 5 percent of the territory of a single historical case of military occupation, the overall contribution of FM’s study—as is typical with DBI research—hinges on its internal validity. The validity of their findings depends on the credibility of the particular claims on which their natural experiment is based: that the placement of the LoD at the local level was indeed as good as random; that the level of devolution of political power to local elites was indeed greater in Vichy France; and that the patterns of resistance in direct proximity of the LoD were not affected by an undetected imbalance between the territory on either side of the LoD.

As we will demonstrate, however, none of these claims is supported by the historical evidence. We reconstruct the process that produced the data FM use. We show that the French Resistance carried out more attacks near the LoD in the directly-occupied zone because the double-track railways used intensively by the Germans to move troops and materiel were more abundant there than they were in the Vichy zone. We explain how the pattern of resistance in this area was shaped by Allied efforts to coordinate Resistance action with their effort to liberate France. We reveal how the LoD was locally delineated by German leaders with the goal of keeping these important railways under direct German control. Finally, we cast doubt on the claim that German rule was more “devolved” to the French in the Vichy zone than it was in the directly occupied zone during the period when most acts of violent resistance occurred. What emerges is an idiographic account of what caused the observed patterns of resistance in these particular fragments of French territory—an account that is consistent with the existing historiography as well as with original macro-level data we introduce.

The remainder of this article proceeds as follows. We first outline the relationship between nomothetic and idiographic explanations in DBI, and we highlight the crucial role of macro-historical evidence in validating micro-level research. We then apply these insights to the subject of violent resistance against the occupation of France during WWII. We begin by providing a brief

sketch of the historiographical problem of estimating the effect of political motivations on armed resistance activity. We introduce a new observational dataset that provides some aggregated evidence about the role of pre-war political ideology and its interaction with the LoD. We discuss the limitations of traditional observational research and the promise of FM’s DBI approach. Next we explore the causes of variation in the level of violent resistance in the fragments of French territory examined in FM’s study. We conclude by discussing the implications of our argument for the conduct of political research and by suggesting evidentiary standards to avoid the problems we highlight.

Experiments, History, and Causal Inference

Causal inference—determining *why* things happen as they do—is a broadly shared goal of political methodology. Most important outcomes in politics result from several causes. A central challenge of political research is to determine the relative weight, or “causal effect,” of each factor. While genuine causes produce statistical associations among variables, the high dimensionality of social life gives rise to many associations that either do not reflect genuine causal effects or do not reflect the true magnitude of the causal effects that in fact obtain. This problem makes advocates of DBI deeply skeptical of the credibility of causal inferences drawn from purely “observational” data. In response, DBI shifts the inferential burden away from the researcher’s ability to correctly model the causal process that produced historical data and toward the active design of research that can isolate and measure the effect of a single factor on an important outcome.

The epistemic ideal of DBI is the RCT, in which the researcher tests a nomothetic proposition about the relationship between cause and effect by analyzing small subsets of a population, organized into control and treatment groups, and then extrapolating from this sample to the broader world. Random assignment guarantees that, in expectation, the groups assigned to each experimental condition will be homogeneous in regard to *all* their pre-treatment characteristics. As a result, the differences we observe among the two groups on an outcome of interest cannot be attributed to systematic differences in the units that existed prior to treatment. Moreover, the precise nature of the treatment is known, because the researcher herself designed it as a means to test a specific nomothetic conjecture. In laboratory and field experiments, the researcher has direct control over the data-generating process, which is documented in a qualitative description of the experimental protocol.

In natural experiments, however, the researcher has no control over the processes of treatment assignment and administration. Instead, natural experiments harness natural or social processes to test a nomothetic proposition.

To ensure that the test is indeed a natural experiment—not “just” an observational test, to which experimentalists would devote a lower level of epistemic trust—the researcher must validate that these historical processes conform to the assumptions of the experimental framework.

To begin with, researchers must validate the process of assignment to treatment, which is, as Dunning points out, the “Achilles heel” of natural experiments.⁹ The existing literature acknowledges two ways of validating the assumption of as-if random assignment. First, researchers can deploy what Dunning calls “treatment-assignment causal process observations” (CPOs).¹⁰ These are qualitative “pieces or nuggets of information about the process by which units were assigned to treatment and control conditions.”¹¹ Second, researchers can conduct quantitative balance tests on relevant pretreatment covariates.¹² These two validation techniques are seen as complementary and largely fungible.

Ideally, the researcher has knowledge of the causal process through which units were assigned to treatment and control groups and can validate the as-if random assumption using qualitative treatment-assignment CPOs. As Hyde argues, “the burden in natural experiments rests on the researcher to provide evidence that the treatment can, in fact, be treated ‘as-if’ it had been randomly assigned.”¹³ Knowledge of these CPOs requires “fine-grained knowledge about context and process”¹⁴ without which “analysts may be studying something less than a natural experiment, and causal inferences drawn from the study may be more tenuous.”¹⁵ Specifically, Dunning recommends that researchers look at evidence “on the information, incentives, and capacities of key actors with control over processes of treatment assignment.”¹⁶

Once a researcher establishes that the process of treatment assignment was as good as random, quantitative balance tests on pre-treatment covariates can be used to assess the possibility that the treatment and control groups are heterogeneous by chance on any *relevant* characteristics, in which case any difference observed in the outcome might derive from these unbalanced characteristics rather than from the treatment.

Sometimes, however, researchers are unable to reconstruct the process of treatment assignment. When this happens, it is tempting to take absence of clear evidence against the as-if random assumption, along with statistical tests showing balance on a set of observable pre-treatment covariates, as sufficient to validate a natural experiment.

We believe this second approach is misguided. Absent compelling qualitative evidence that the causal process that assigned units to treatment and control groups was, in fact, as good as random, natural experiments should not be accepted as valid. Specifically, without convincing treatment-assignment CPOs, the as-if random assumption cannot be validated solely using quantitative balance tests. If this were possible, then any sample of observational data

could be considered a natural experiment as long as treatment and control groups were balanced on all the variables a researcher has observed. But this is clearly not the case. Experimentalists in political science object to observational studies precisely because controlling for observables can never rule out the possibility that an estimated causal effect is the artifact of some unmeasured confound. The reason is simple: without knowing the causal process through which units were assigned to treatment and control groups, researchers cannot determine the set of *relevant* confounds on which it is important that the two groups be balanced. In this scenario, the only reliable alternative would be “to measure each and every one of the unmeasured factors. The intrepid researcher who embarks on this daunting task confronts a fundamental problem: no one can be sure what the set of unmeasured factors comprises. The list of all potential confounders is essentially a bottomless pit, and the search has no well-defined stopping rule.”¹⁷ In sum, quantitative balance tests, no matter how numerous, are no substitute for reliable qualitative evidence of the causal process that produced the data. The two are not fungible. Treatment-assignment CPOs enjoy epistemic priority over balance tests.¹⁸

Validating a natural experiment also requires that the researcher reconstruct the process through which the treatment was applied, so as to ensure that the data tests the general nomothetic proposition being evaluated. This entails two additional types of qualitative historical data: “independent-variable CPOs” and “mechanism CPOs.”¹⁹ Taken together these two types of idiographic account establish the chain connecting cause and effect in the data. The first step in reconstructing this chain is to collect independent-variable CPOs, “nuggets of information . . . about the presence or values of an independent variable (a treatment).”²⁰ No natural experiment should be considered valid without a qualitative account of “whether a cause occurred in the manner and/or at the time posited by the theory.”²¹ This entails gathering historical knowledge specifying what exactly is the treatment.

Furthermore, validating a natural experiment requires the researcher to collect mechanism CPOs, which “provide information not just about whether an intervening event posited by a theory is present or absent but also about the kinds of causal *processes* that may produce an observed effect.”²² Any nomothetic proposition connecting cause and effect presupposes a causal chain of variable length linking the two. Particularly when any causal process compatible with the theoretical proposition being tested is complex and operates over long spans of time, researchers must process-trace the pathway from the treatment to the outcome. This entails gathering historical knowledge on the chain of events through which the treatment leads to a different outcome than would have occurred in its absence. Without knowledge of this

process, experimental results cannot be taken as evidence in support of a particular theoretical proposition. Qualitative research on the process of treatment administration also cannot in principle be replaced by however many quantitative outcome tests. While independent-variable and mechanism CPOs are seen as merely desirable in the existing literature, we argue that they are necessary to validate a natural experimental test of a nomothetic proposition.

Finally, we argue that when reconstructing the historical processes of treatment assignment and administration, researchers must strive to avoid the danger of reductionism, to which natural experiments are particularly likely to fall prey, given their typically micro level of analysis. To build qualitative accounts of the treatment assignment and administration processes, researchers must uncover the intentions, deliberations, and decision-making procedures of the people whose actions the data record. This requires attention to multiple levels of analysis, from the micro-level at which experimental research is typically conducted, to the macro-level historical dynamics shaping the data. The level of spatial or temporal aggregation at which crucial evidence about how the data were produced may be located cannot be pre-determined. Therefore, while natural experiments are typically run on highly disaggregated micro-level data, researchers cannot safely ignore the macro-historical context in which the data are situated. If the researcher misunderstands the context in which the data were generated, causal inferences based on natural-experimental results may be unwarranted.

Taking stock, historical process tracing is not just a useful methodology when validating natural experiments; it is a requirement, since the epistemic trust placed in natural-experimental results depends on the accuracy of the underlying idiographic accounts of the treatment assignment and administration processes. As Collier, Brady, and Seawright put it, good qualitative evidence is an essential component of successful quantitative analysis and, therefore, of the quest to improve our standards of causal inference.²³

Put differently, we maintain that the validity of natural experiments as a method for testing general causal propositions depends logically on the truth status of causal statements about unique historical processes and events. If, however, credible causal inferences can be drawn from qualitative historical data when the assumptions of a natural experiment are at stake, one may reasonably ask why idiographic causal inference should be regarded as, in general, unfeasible, unscientific, or otherwise inferior to nomothetic causal inference based on natural experiments. Either one trusts historical research, and natural experiments are valid but only one of several modes of studying politics; or one does not trust historical research, and natural experiments are not an acceptable tool for generating scientific knowledge.

Exactly *how* researchers should approach the problem of causal inference for specific units in such cases is beyond the scope of this paper. In this sense, we provide a conditional rather than a direct argument: if a natural experiment is possible, then it necessarily depends on one or more prior causal inferences about unique cases, which cannot themselves be validated by means of an experiment. It may seem like this requirement sharply distinguishes putative natural experiments from “true” randomized experiments. But the credibility of any given RCT also depends on the fidelity of the unique causal process generating the experimental conditions to the specifications and assumptions laid out in the experimental protocol. From the point of view of a third party—i.e., not the experimenter herself—validating the assumptions of a natural experiment may be more challenging than evaluating a contemporary RCT, but the difference is one of degree rather than one of kind.

Resistance to Foreign Occupation in World War II France

We now provide an empirical illustration of the critique we advanced in the previous section by examining Ferwerda and Miller’s recently published natural experiment on the French Resistance.²⁴ FM examine how an institutional difference (German devolution of authority to the Vichy government) interacted with political partisanship to cause variation in violent resistance. We first examine what the general historical literature can tell us about this question. Given that FM’s study focuses on a small fraction of French territory, we next look at what aggregated data covering the whole of France says about the association between partisanship and armed resistance. Finally, we consider the limitations of observational data, and the attractiveness of FM’s approach to establishing causality in this case and beyond.

*Partisanship and Resistance in the Historical Literature*²⁵

Individual reactions to the German occupation of France varied tremendously, from outright enthusiasm to dogged resistance. The overwhelming majority of Frenchmen disliked the occupation but did nothing concrete to contest it. Even “moral resistance” was not common; organized, armed resistance was exceedingly rare for most of the period. The historical literature stresses the ideological and political diversity of the Resistance, which included conservative military officers in the mold of Charles de Gaulle alongside the Comintern-directed French Communist Party (PCF), and everything in between.²⁶ With the exception of the PCF, the Resistance had no explicit basis in pre-war party politics.²⁷

Nevertheless, the collaborationist Vichy regime was unmistakably right-wing, drawing the majority of its personnel and active supporters from conservative

segments of society: military officers, local “notables,” and the activists of pre-war conservative parties and factions. The authoritarian tools furnished by the dissolution of the Third Republic permitted Vichy to halt and reverse the long leftward drift of French politics during the interwar period, and in particular to redress the effects of the 1936 election, which had brought the left-wing Popular Front coalition to power. Vichy launched a “National Revolution,” ostensibly to revitalize France through patriotism, corporatism, and a return to traditional social values. The regime persecuted the political left in a variety of ways, interning and sometimes deporting suspected communists, purging many left-leaning civil servants and local elected officials, and putting key figures of the Popular Front government on trial.

The character and policies of the Vichy regime gave Frenchmen on the political right good reasons to support it and to acquiesce toward the presence of their German allies. Conservatives feared the return of the French left to power, especially in a revolutionary form. Building on this mindset, German and Vichy propaganda worked hard to convince the French that the victory of the Resistance would mean the victory of Bolshevism.²⁸ French citizens on the political left had corresponding motives to support the Resistance, as the depth of Vichy’s collaboration with Germany and the extent of its ideological partisanship became increasingly apparent over time.

It is a matter of scholarly consensus that markedly different conditions prevailed *initially* on opposite sides of the LoD. The origins and developmental trajectory of the Resistance in the two zones were profoundly shaped by these differences. Repression was more effective and severe in the occupied zone, which made it more difficult to organize resistance there. As a consequence, the two zones tended toward distinct organizational forms. Small, clandestine, and often specialized “networks” were more common in the occupied zone, while larger “movements” with a significant public dimension arose in the Vichy zone.²⁹

While the *opportunity* to resist may have been more limited under direct German occupation, FM argue that right-wingers in the Vichy zone were less motivated than their homologues in occupied territory to oppose arrangements that empowered their political faction. This interpretation was advanced long ago by the French historian Henri Michel, who pointed out that the largest and best-known resistance movements in the unoccupied zone—*Combat*, *Libération Sud*, and *Franc-Tireur*—all had a left-wing orientation, while the many networks in the occupied zone were more ideologically diverse.³⁰

While the argument advanced by Michel and tested by FM is plausible, several things cut against its relevance for explaining *violent* resistance in particular. First, although it is possible that French citizens were motivated solely by the political context within their own occupation zones, the nationwide political context could have mattered just as

much or even more to them. Even locally, the victory of the Resistance had the potential to threaten the interests of right-wingers anywhere in France. Second, the Vichy zone was unoccupied only until November 1942, when the Germans invaded the south and Germany’s devolution of authority was largely reversed. Shortly after, the LoD that had divided the two zones was decommissioned. Third, and most importantly, nearly all of the violent resistance to the occupation occurred during 1943 and 1944, after the governing institutions of the two zones had substantially converged.

Partisanship and Resistance in Aggregated Data

What evidence do we have about the relationship between partisanship and violent resistance in WWII France? Given the loose connection between pre-war party politics and the Resistance, it is difficult to arrive at a synoptic picture of how partisanship and the institutional differences between the occupation zones influenced individual motivations to resist with solely qualitative methods. We examine here the association between partisanship and violent resistance across all of France using an original dataset aggregated to the level of the department ($N = 86$). We control statistically for a number of factors that might have been associated with partisanship and resistance. Our dependent variable, coded from primary sources held by the French National Archives, is an event count for 1943–1944 of railroad sabotage, one of the main forms of violent resistance in wartime France.³¹

We coded two indicators, one for departments entirely in the Vichy zone (*Vichy*) and a second for the thirteen departments intersected by the LoD (*Line*); the occupied zone is the implicit third category. To measure partisanship, we consulted returns from the first round of the 1936 parliamentary elections. Since the party system of the late Third Republic was highly fragmented, we followed Lachapelle’s classification of ideological groupings and calculated the 1936 vote share for both the left and right.³² These two variables are highly collinear, so we use the vote share for the right minus the share for the left (*Vote share difference*). Given that votes not cast for the right (left) might go to centrist parties rather than the left (right), we control for the vote share of moderate parties (*Vote share center*). We also control for the log of department population (*Log population*), taken from the 1936 census. Railroads were not evenly distributed across France, so it would make sense to find more railroad sabotage where there are more railroads. We therefore control for the length of double-track line (the type of railroad most often attacked by the Resistance) in each department (*Double track length*).

We include a variable for the percentage of each department’s land area that is mountainous (*Rough terrain*). Mapping the data suggests a clear pattern of higher sabotage counts in proximity to Germany, so we include the distance of each department’s centroid to the German border to

capture this pattern (*Germany distance*). To capture the possibility that partisanship affected (i.e., dampened) resistance more in the Vichy zone, we interact *Vichy* and *Line* with *Vote share difference*. We fit a poisson regression to the data. We report our estimates in table 1. (Appendix A includes descriptive statistics on the variables in table A.1.).

Figure 1 plots (from left to right) the predicted sabotage count ranging from departments dominated by the political left to departments where the right was stronger in the 1936 elections, separately for the occupied zone and the Vichy zone, with all other variables at their means.

We find that left-leaning departments were strongly associated with higher levels of violent resistance throughout France. The slope of the relationship between partisanship and sabotage is roughly similar for the two zones, with tight confidence bands, suggesting that the left-right cleavage was associated with resistance for France as a whole.³³ The predicted counts of sabotage for the Vichy zone are higher across the board. Furthermore, we find that the presence of double-track railways was, unsurprisingly, associated with the incidence of sabotage attacks. In sum, our sub-national data on WWII France is consistent with the view that political partisanship shaped the motivation to resist in both the occupied and Vichy zones in a roughly similar way. We find no

evidence in these data to suggest that this association was different on the two sides of the LoD.

The Line of Demarcation and Resistance in World War II France

The quantitative data presented in the previous subsection, while perhaps not strictly useless, is defective in an obvious manner: there is no way to rule out the existence of an unmeasured variable that is correlated with both partisanship and armed resistance. In the DBI language, there is no “identification strategy”—no strategy to identify the causal effect of partisanship on armed resistance. Even if political attitudes really did cause variation in violent resistance in a similar way on both sides of the LoD, our approach would not permit a reliable estimate of the magnitude of this effect.

For both WWII France and the more general topic of military occupation, these concerns are not merely theoretical. Foreign powers are likely to base their decisions about which regions to occupy, in part, upon the pre-war characteristics of those regions. Even when they do not, regional variation within countries on a single variable of interest—for example, partisanship—will almost always be associated with variation on a number of other potentially influential variables, not all of which can be measured easily.

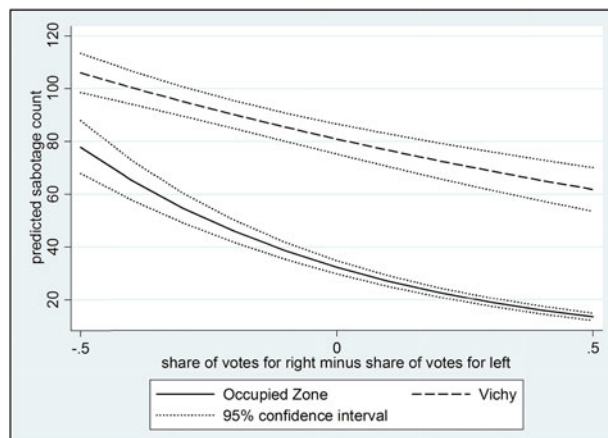
To address these shortcomings of traditional observational research, the FM study deploys multiple tools in the DBI kit. FM start by carefully selecting a subset of the data in which the many potentially confounding characteristics of French political units can be controlled “by design.” Unlike any other country occupied in WWII, France was split into occupied and unoccupied zones by a fixed boundary rather than a moving military front line.

Table 1
Determinants of sabotage in wartime France

	Poisson model
<i>On Line</i>	1.024*** (.046)
<i>Vichy</i>	0.917*** (.063)
<i>Vote share difference</i>	-1.757*** (.089)
<i>On line X vote share difference</i>	1.362*** (0.132)
<i>Vichy X vote share difference</i>	1.218*** (0.119)
<i>Vote share center</i>	0.389* (0.115)
<i>Log population</i>	0.062 (0.033)
<i>Germany distance (km)</i>	-0.003*** (0.0001)
<i>Rough terrain (% of land area)</i>	-0.007*** (.0009)
<i>Double track length (km)</i>	0.002*** (0.0002)
Constant	3.804*** (0.397)
Pseudo-R ²	0.596
Observations	86

*p<.05, **p<.01, ***p<.001

Figure 1
Marginal effect of partisanship on sabotage in wartime France



Moreover, the LoD was imposed by the Germans and did not follow any pre-existing political divisions, which might have corresponded to other important pre-existing differences. As a result, we might expect variation to be “smooth” around the discontinuity created by the LoD and therefore well suited to the regression discontinuity design (RDD) FM employ. Additionally, it is indisputable that the occupied and Vichy zones were initially subjected to different policies and, consequently, had distinctive wartime trajectories. Finally, FM focus on a dependent variable—sabotage—on which the French case offers great sub-national variation. In sum, FM’s research design configures a nearly ideal implementation of the DBI guidelines, leading them to claim that their analysis is “perhaps the first causally identified evidence that extending political authority to natives can reduce levels of resistance.”³⁴

Whereas we find no interaction between partisanship and the LoD, FM find that armed resistance was more common in the occupied zone, which they attribute to the demobilizing effect of indirect rule on right-wing Frenchmen in Vichy. Should we discard our argument on the equally partisan character of the French resistance, loosely supported by the evidence we presented on the entirety of French metropolitan territory, and instead adopt FM’s argument on the differential partisanship of the Resistance across the LoD, based on evidence collected in less than five percent of France?

To adjudicate this question, in the next section we respond to Dunning’s call for scholarly “adversarialism” in the validation of natural experiments and zoom in on the four departments used in FM’s study.³⁵ By reconstructing the historical process that generated violent resistance in this fragment of French territory, we lay out an alternative causal account of the variation in the level of resistance across the LoD in its immediate vicinity. In doing so, we reveal how this section of French territory is inadequate to test general propositions about the effects of political devolution on resistance to foreign occupation—or, for that matter, to test any other nomothetic propositions by postulating that the local placement of the Line of Demarcation configures a natural experiment.

Strategic Logic, Political Rule, and Resistance in World War II France

In this section, we reconstruct the historical process through which violent resistance was produced in the territory that FM analyze: the departments of Charente, Cher, Saône-et-Loire, and Vienne. We lay out an alternative causal account that explains variation in the sub-national pattern of violent resistance near the LoD as a function of the interaction of German and Allied war strategies. We substantiate each element of our argument using qualitative historical evidence. We show how the pieces of our argument fit together using detailed maps. We conclude

with sharp quantitative tests of several elements of our explanation.

While our evidence is complex and detailed, our argument is simple. Near the LoD, the Resistance staged more sabotage attacks in the occupied zone than in the Vichy zone because the targets they wanted to attack—the double-track trunk lines of the French railroad system—were more plentiful in the occupied zone. When the overwhelming majority of these attacks occurred, the LoD was not a meaningful obstacle to the mobility of the Resistance. Furthermore, the Resistance had by then created a national command structure, which coordinated with the Allies to attack railways in order to impede the movement of German troops and materiel toward the sites of the Allied invasions and, later, away from the collapsing front lines. Consequently, Resistance groups based on one side of the LoD could easily strike targets on the other and, given the greater prevalence of important railways on the occupied side, obeyed a strategic imperative to do so. Finally, these railroads were more common in the occupied zone because the LoD was *not* locally placed as-if randomly. The Germans chose its path in 1940 to keep these major railways under their direct control. At the same time, the governing institutions that prevailed on the two sides of the LoD during this period were not appreciably different, making the French case a poor test of the effects of indirect rule. In other words, we uncover a case of spurious causation: the location of important railways—a “pre-treatment” variable—explains the variation of both the supposed cause and the outcome.

The Strategic Logic of Violent Resistance in World War II France

A central aspect of our argument is that both the Resistance and the environment it operated in changed over time. Therefore, we begin by looking at the timeline of violent resistance. To do so, we reconstructed the dataset of violent events in FM’s four departments from its original source, the departmental histories of the Resistance compiled by the French *Service Historique de l’Armée*.³⁶

Figure 2 depicts the monthly incidence of sabotage from June 1940 through September 1944. As one might expect from a gradually mobilizing resistance movement, the rate of attacks increased over the course of the occupation. The rate of increase, however, exhibits two marked discontinuities, in April 1943 and June 1944. Table 2 uses these discontinuities to split the data into three periods.

During the first 34 months of the occupation, when differences in the structure of authority between the occupied and unoccupied zones were starkest, sabotage was virtually non-existent. This period accounts for two-thirds of the occupation (34 out of 51 months) but less than 5 percent (32) of the sabotage.³⁷ When one part of France was occupied and another was not, the rate of

sabotage did not vary across the two zones because it was a constant at nearly zero. In effect, when FM's theory would lead us to expect particularly strong differences in *motivations* to resist on opposite sides of the LoD, there is virtually no violent resistance and, thus, no variation to study.

This situation changed markedly during 1943, as violent resistance picked up steam; the monthly attack rate in the subsequent period, from April 1943 through June 5, 1944, was approximately 22 times greater, accounting for an additional 41 percent (285) of the sabotage events. This increase followed a highly consequential event: the German invasion of the Vichy zone in November 1942 and the abolition of the LoD in March 1943.³⁸

From its implementation in 1940 until its abolition in 1943, the LoD was a very real obstacle to movement

and communications, with physical barriers and checkpoints. After March 1943, crossing the LoD was “a mere formality.”³⁹ The checkpoints and customs personnel that had regulated movement between zones were withdrawn.⁴⁰ Thenceforth, the LoD was a legal abstraction, and no particular hurdles precluded a resistance group based in one zone from conducting actions in the other.

The absence of regulation at the boundary creates a risk of “spillover” or “interference,” a situation that arises when a treatment delivered to one unit affects the outcome of one or more additional units. Because an RDD focuses on a small strip of territory along the LoD, communes that were under direct German occupation for the entire war often lie within easy walking distance of communes in Vichy France. Since Resistance groups were mobile, any policy applied on one side of the boundary could have affected sabotage on the other side. After March 1943, armed groups based in the *zone nord* could have carried out attacks in the *zone sud*, and vice versa, if they had had good reasons to do so—precisely what we will demonstrate.

The problem of spillover is compounded by a second factor: the growing centralization of resistance efforts. Although the initial impulse to mobilize against the Germans was often local, as the occupation dragged on the Resistance evolved toward a unified structure. During the first half of 1943, Charles de Gaulle's representative Jean Moulin brought together the leaders of the most influential Resistance groups on both sides of the LoD, including the Communists, under the aegis of the National Resistance Council (*Conseil National de la Résistance*), which “formally acknowledged the leadership of de Gaulle.”⁴¹ Beginning in August 1943, the secret services of the Free French created a network of regional military delegates, whose mission was “to organize and coordinate the military resistance of the movements” at a regional level.⁴² From this point on, the Resistance possessed a structure capable of organizing violent action across localities and even regions. Over 95 percent of the sabotage events in these four departments took place once

Figure 2
Chronology of Sabotage data per department (Charente, Cher, Saône-et-Loire, and Vienne), June 1940–September 1944

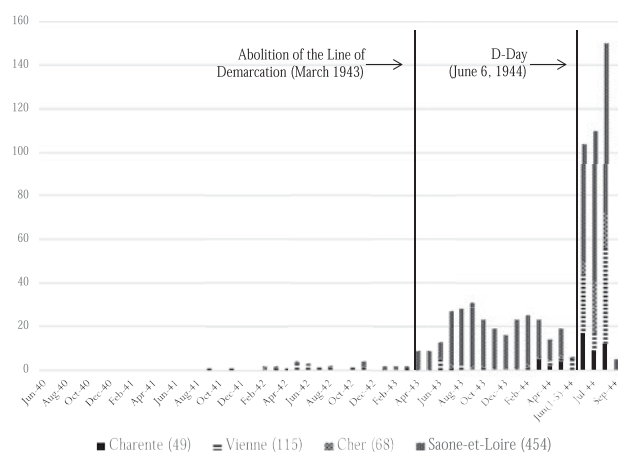


Table 2
Distribution of Sabotage per department per period

	June 1940–March 1943	April 1943–June 5 1944	June 6–September 1944	Total	% of Total
Charente	0	11	38	49	7.1
Vienne	18	18	79	115	16.8
Cher	7	15	46	68	9.9
Saône-et-Loire	7	241	206	454	66.2
Total	32	285	369	686	100.0
% of Total	4.7	41.5	53.8	100.0	
Monthly Rate	0.9	20.1	96.3	13.2	

Note: The data include 46 other events without a complete date. Of these, 35 are coded as having taken place in 1944, after the abolition of the Line of Demarcation.

this process of consolidation had started in early 1943. Although Resistance groups continued exhibiting differences during this later phase of the occupation, they had a higher degree of coordination, which made them capable of operating beyond their local base as part of an integrated effort to liberate France.

Taken together, the abstract nature of the LoD post-1943 and the possibility of spillover between the two zones should lead us to question the use of sabotage data in these four departments to test the validity of arguments that infer the geographic origin of the perpetrators in relation to the LoD from the precise location of sabotage attacks.

But perhaps the most problematic dynamic at play in this period from the perspective of using these data to test a theory about the effects of political rule is the diluted difference between the political regimes of the zones *nord* and *sud* once all of France was occupied. It is true that an administrative division was maintained, but this was far from being the same administrative division that had preceded the occupation of the *zone sud*.

Nominally, the Vichy government continued to exercise authority on both sides of the LoD, and most aspects of the day-to-day administration of the two zones were carried out by French officials. Rather than place the *zone sud* under the pre-existing Military Administration of France (*Militärverwaltung Frankreich*), the Germans created a parallel structure directly under the Western Front commander.⁴³ Despite a distinctive chain of command, the structure of authority in the two zones converged in many respects. Several German field divisions were cantoned in the *zone sud*; it was, in this sense, a genuine occupation. Vichy's "armistice army" of 100,000 men was dissolved. German internal security services, including the Sicherheitsdienst (SD) and the Gestapo, began operating in the *zone sud*.⁴⁴ Teams of military liaisons were placed in its regional and departmental prefectures and sub-prefectures. According to Jäckel, these teams "roughly corresponded, without being so named, to the Field and District Commanders (*Feld- und Kreiskommandaturen*) of the occupied zone; they also had units of the [German] military police (*Feldgendarmarie*) under their orders."⁴⁵

Accordingly, we find little support in the historiography for the view that "the extent of native authority continued to differ between the two zones."⁴⁶ Some of the most prominent studies of the period point in the opposite direction. Paxton notes that "the practical conditions of life in the Vichy zone were now little different from those in the north."⁴⁷ Jäckel refers to the French government as having a "fictitious existence" after January 1943, noting that it "had lost all the means of pressure it had previously used in talks with Germany."⁴⁸ From November 1943 on, Jackson describes Vichy as "a State which barely existed."⁴⁹

Our point is not that the Germans administered all of France directly after November 1942. Given their heavy reliance on French administrators, it could be argued that no part of France was under direct rule during the war. Gildea describes the "crucial interface" between the Field Commanders and the Prefects, their French counterparts *in the occupied zone*, as similar to "a system of 'indirect rule' as developed in the British Empire by Frederick Lugard, former governor-general of Nigeria."⁵⁰ What matters is the absence of clear evidence that power was devolved to a greater degree in the *zone sud* than in the *zone nord* during the final 22 months of the occupation.

Using the language of experiments, FM would have us think of native authority as a discrete "treatment" with varying "doses," such that the *zone sud* received a high dose initially, and then a lower dose following the German occupation of the region. The causal effect is treated as "cumulative," capturing the effect of the high initial dose and the subsequent lower dose. But it is difficult to say whether the policies applied in the two zones after November 1942 differed in the amount of authority devolved to the French, and if so by how much.

Alternatively, one might think of the *zone sud* as having been sequentially "treated" with two distinct policies, the second of which was very similar to the policy implemented in the *zone nord* for the entire war.⁵¹ This convergence of policies raises the question of exactly what treatment was actually received in the *zone sud* after November 1942 and how it differed from the *zone nord* "control" group. If both zones were under a common policy when over 95% of the sabotage occurred, then these data do not allow for a valid test of the effect of indirect rule. The evolution of German governance highlights a way in which qualitative methods are essential for validating natural experiments: by deploying independent variable CPOs and mechanism CPOs to ascertain whether or not the posited treatment is actually delivered and to what extent it differs from what the control units receive.

The second discontinuity in the chronology of violent resistance occurs on June 6, 1944, when the monthly sabotage rate jumped almost five-fold over the previous period (and over one hundred-fold over the initial period), reaching an average of 96 attacks per month in just these four departments until September, when their liberation was concluded. This period of slightly less than four months accounts for the majority (53.8 percent, or 369) of all the sabotage attacks.

As is well known, on June 6, 1944, Allied forces invaded Normandy, reopening conventional land warfare in France. After a two-month stalemate, the German front collapsed in the second week of August, shortly before the Allies invaded the Mediterranean coast on August 15. These two developments led to the ejection of German armies from nearly all of France by the middle of September.

D-Day had at least two important effects on the pattern of violence. First, the prospect of liberation changed the political calculus for both the Resistance and the many officials who had collaborated with the Germans. Authority was destabilized throughout France, making it difficult to pinpoint the type of rule that prevailed in each location at each moment. For instance, Gildea writes: "By the summer of 1944 the Germans had only limited cooperation from French officials in the regions, who were preparing to switch their allegiance either to the Allied forces when they appeared or to de Gaulle's provisional government."⁵² Jackson shows that even Pierre Laval and Philippe Pétain, the two highest-ranking figures of the Vichy government, put schemes in motion "in the hope of presiding over the transition between Vichy and the post-Liberation regime."⁵³ The Resistance saw opportunities for insurrections ahead of the advancing allied forces. While some uprisings were launched autonomously, others were facilitated or even directed by commandos parachuted into France by the Allies. The result was fragmented authority on both sides of the LoD, rather than two discrete zones with distinctive authority structures.

Second, after D-Day armed resistance was explicitly coordinated with the Allies. Beginning in late 1943, the British Special Operations Executive (S.O.E.) and the Free French *Bureau Central de Renseignement et d'Action* devised a plan to hinder railway traffic in France after the landings through a combination of aerial bombing and sabotage by the Resistance.⁵⁴ This effort led to *Plan Grenouille*, aimed at sabotaging railway depots, and *Plan Vert*, aimed at sabotaging railway lines.⁵⁵ Their purpose was to delay the transfer of reserves to the Atlantic coast following D-Day.⁵⁶ To this effect, "containers marked 'Plan Vert' were parachuted into each region to be allocated according predefined priorities. . . . Each [French Resistance] team received one copy of a map . . . containing the approximate location of the point to be attacked."⁵⁷ Given the limited means of the Resistance, "only the most important itineraries, given the Allies' knowledge of the enemy's order of battle, were selected."⁵⁸

The fluidity of authority and the importance of coordination by the Allies can be seen in greater granularity by turning to the specific events in the Vienne and the Saône-et-Loire, which together account for 83 percent of the total sabotage attacks in these four departments. In the Vienne, the Resistance emerged only in the spring of 1944 "in anticipation of the approaching landing" of Allied forces.⁵⁹ Starting in May, the Allies parachuted over 4,000 containers with weaponry and ammunition.⁶⁰ In exchange for these supplies, local Resistance groups (integrated across the former LoD) agreed to be placed under the unified command of Major Maingard de la Ville, himself part of a chain of command ultimately directed by General

Koenig, de Gaulle's delegate to the Allied headquarters and, after D-Day, commander of the French Forces of the Interior [FFI].⁶¹ To increase the Resistance's combat strength, "'Jedburgh' teams, composed of three men (one American, one English, one French) were sent in to help familiarize the new combatants in weapons manipulation and guerrilla techniques. On the ground, the FFI received reinforcements from the [British] SAS (Special Air Service): 56 men were parachuted in between June 6 and 10, 1944, [and] four Jeeps arrived by air." Their purpose was "to multiply the sabotage" against major transportation axes, "with the goal of delaying as much as possible the routing of [German] troops, materiel, and munitions."⁶²

Increasing the strategic importance of double-track railways in the Vienne, the August 15 landings on the Mediterranean coast required German armies stationed in southwestern France to redeploy to Dijon through Tours and Bourges (refer to map 1). As Calmon writes, it was "a matter of escaping the two jaws that are closing" as Allied forces moved East from Normandy and North from Provence.⁶³ Overall, then, the Resistance in the Vienne now acted as a behind-the-lines force supporting these Allied offensives.

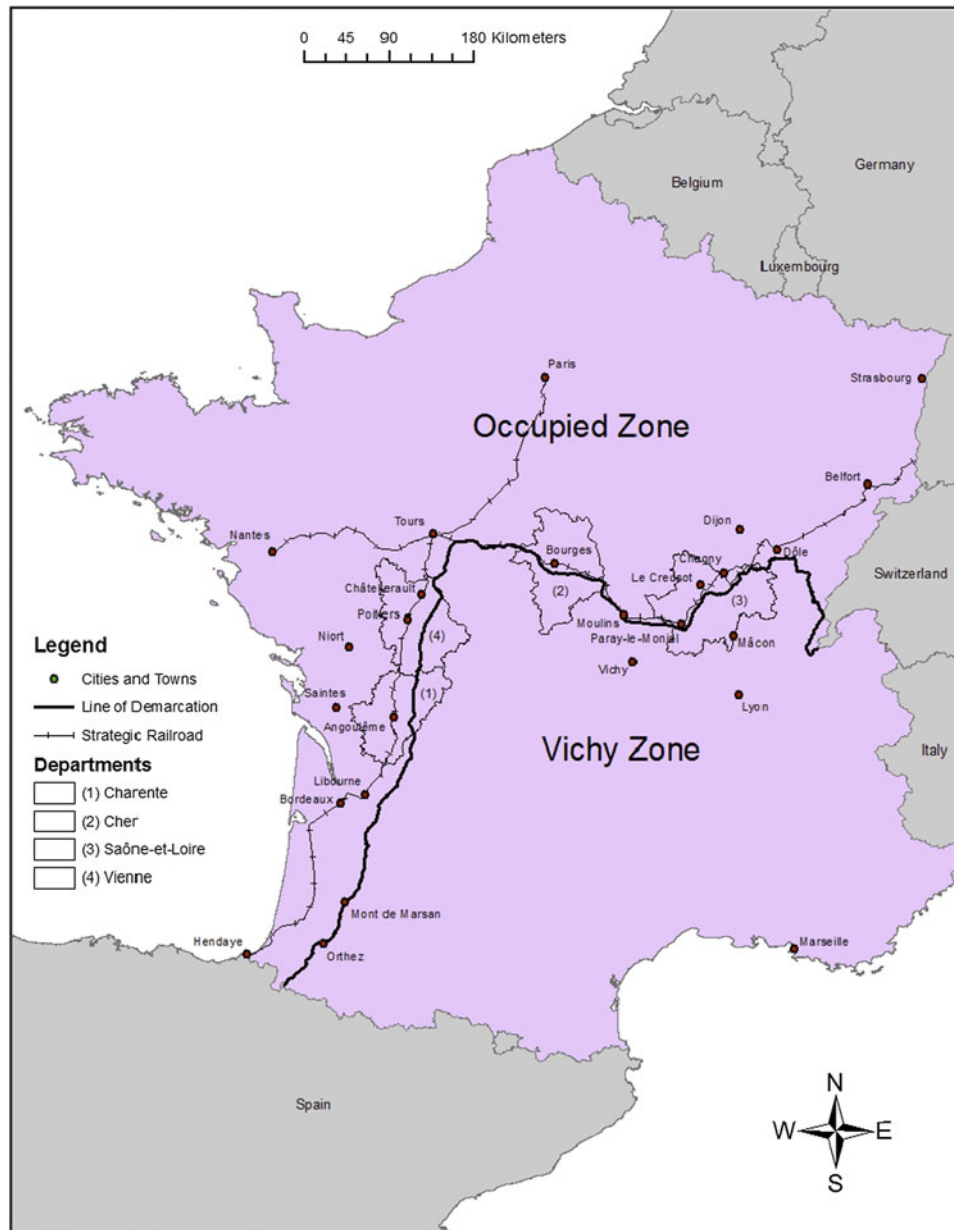
The situation was similar in the Saône-et-Loire, which accounts for 66 percent of all sabotage events. The Saône-et-Loire was crisscrossed by some of the most important railways in France, linking the Mediterranean coast, through Lyon, to both Paris and Germany, and connecting to a major East–West route that terminated near the Allied front in Normandy.⁶⁴ Given its strategic importance, the British S.O.E. supplied the Resistance there with thousands of weapons and explosives before the invasion.⁶⁵ Soon after D-Day, the Resistance liberated a triangle of territory straddling the LoD.⁶⁶ From this base, Resistance operatives launched repeated raids against the Paris-Lyon-Marseille and Strasbourg-Bordeaux railway lines *on both sides of the LoD*.⁶⁷

Impeding railway communications became even more important following the August 15 landings on the Mediterranean coast. To help prevent the escape of Wehrmacht units, the S.A.S. parachuted teams of saboteurs into the Cluny area on August 13 and 14, 1944. Their objective was to "accelerate the rhythm of the operations against the communication lines taken by the retreating German forces."⁶⁸ According to a French military commander, "the withdrawal of the 1st German army to Dijon, ordered before the Allied landings in the south of France, was made almost impossible by the actions of the F.F.I. and there too the sabotage of the railway network played a decisive role; only a quarter of the German forces stationed in the South-West succeeded in reaching Dijon."⁶⁹

Taking stock, for these data to constitute a test of a theory on the effects of differences in political rule across the LoD, violent resistance would have had to be the

Map 1

The Line of Demarcation (LoD) and Nantes-Tours-Belfort and Paris-Tours-Bordeaux railroads



result of a causal pathway leading from political institutions through political preferences to violence within the extremely local horizons of French communes. A great deal of qualitative evidence suggests that this was not the process that produced most of the events in the data.

Instead, a majority of the sabotage was part of the Allies' strategic plan to liberate France. Within this context, disabling the railways used to transport troops and supplies over long distances was a central objective for

the Resistance, which had undergone a long-term process of consolidation under the authority of the Free French. If we are correct, the large differences between the rates of sabotage in the two zones are driven by a greater availability of major railways as targets in the *zone nord*, not by differences in the structure of political rule.

At this point, one might retort that our critique is irrelevant, given that the LoD was locally delineated in an as-if random manner, so that the causal inferences based

on this empirical design should deserve our confidence notwithstanding the criticisms we presented. After all, natural experiments are designed to produce unbiased causal estimates despite the presence of myriad confounders—such as the strategic variables our criticism emphasizes—in observational data. In the following section, we deploy additional historical data to undermine the claim that the local placement of the LoD constitutes a natural experiment.

The Line of Demarcation: A Natural Experiment?

The key identifying assumption in a natural experiment is the as-if random assignment of units to treatment and control groups. In this case, the assumption is that French communes close to the LoD were “assigned on a quasi-random basis to the Vichy or German zones.”⁷⁰ Although Germany had strategic objectives in the design of the LoD, “most notably, seizing the Atlantic coast and capturing large provincial capitals,”⁷¹ the LoD “followed an arbitrary course at the local level, cutting across preexisting administrative borders and splitting departments (provinces), cantons (counties), and communes (municipalities),”⁷² such that the “motivation behind the precise location of the line remains opaque.”⁷³ If communes were assigned to the occupied and Vichy zones as-if randomly, then the railways the Resistance wished to attack should have been just as common on either side of the LoD. Yet we know that sabotage attacks were much more common on the occupied side. Why?

The documentary record of German decision-making around the time of the 1940 Franco-German Armistice supplies abundant evidence that the LoD was *not* placed as-if randomly at the local level. Rather, a key element of the German rationale for its placement was to keep major railroads running from Germany to the Atlantic coast, and along the coast to the Spanish border, under their direct control. This policy was formulated by Adolph Hitler himself, formalized in the Armistice, and implemented on the ground by the occupation authorities with remarkable faithfulness.

The proposed path of the LoD was discussed on June 17, 1940, in a meeting attended personally by Hitler and two of his top commanders, Generals Alfred Jodl and Wilhelm Keitel.⁷⁴ The location of major railways was a central concern at this meeting:

The envisioned demarcation line between occupied and unoccupied territory was drawn on a map by General Jodl. In the course of doing so, attention was paid to ensuring that the East-West connection through central France that went from Belfort through Dôle-Le Creusot-Moulins-Bourges-Tours to Nantes, and the North-South connection from Tours through Angoulême-Bordeaux to the Spanish border, would run within the territory to be occupied.⁷⁵

Hitler explained that the boundary of the French “sovereign domain” would be determined by the military

necessities of the continuing war with Britain.⁷⁶ Direct control over the major railroads running through *Mittel-frankreich* from Germany to the Atlantic coast and Spain gave the Wehrmacht the mobility to take the fight to Britain or meet emerging seaborne threats.

Indicative of the strategic importance of major railroads, the topic percolated through other meetings Hitler had around this time. He discussed with Mussolini the need to maintain control over a railway link to Spain during their meeting in Munich the following day, June 18. Specifically, “the Führer proceeded to discuss in detail the conditions of an armistice. With the aid of a map, he indicated the form the occupation of French territory should take. . . . In the country’s interior, the occupation would be designed in such a way that under all circumstances the railway leading to Spain via Irun is located completely inside the occupied zone.”⁷⁷ Hitler also insisted on his determination to maintain a railway route to Spain in a meeting with the Spanish General Vigón on June 16, 1940.⁷⁸

Implementing Hitler’s directives, Article II of the Armistice specifies that “French state territory north and west of the line drawn on the attached map will be occupied by German troops.”⁷⁹ The original text of the Armistice, including this map, was retained by Hitler and subsequently lost. However, an Associated Press wire originating in Berlin and containing the text of the Armistice was widely published in international newspapers on June 26 and June 27, 1940.⁸⁰ In lieu of a map, this report included a note on the LoD’s placement that read:

The line mentioned in Article II of the armistice agreement begins in the east on the French-Swiss border at Geneva and runs thence nearly over the villages of Dole, Paray, Le Monial [sic] and Bourges to approximately twenty kilometers (about twelve miles) east of Tours. From there it goes at a distance of twenty kilometers east of the Tours-Angoulême-Libourne [sic] railway line and extends through Mont de Marsan and Orthez to the Spanish border (New York Times, June 26th 1940).⁸¹

This note, published prior to the implementation of the LoD, explicitly situates it relative to the position of the Paris-Bordeaux railway (which ran through Tours, Angoulême, and Libourne), while Dôle, Paray-le-Monial, Bourges, and Tours all lay along the double-track line connecting Tours with Germany, which ended up just inside the German-occupied zone (refer to map 1).

The German rationale for the local position of the line makes eminent strategic sense. Due to a well-documented shortage of trucks, the Wehrmacht relied particularly heavily on rail transportation.⁸² In 1940, France had a very extensive rail network, but not all of the lines were of equal importance. Some carried passengers and freight between economically or strategically important places, while others linked smaller population centers to the national network. The most important lines, its

long-distance “trunk” lines, tended to be double-tracked. Double-track railroads were critical for WWII combatants, because they permit more than twice the throughput of single-track lines.⁸³ The route through central France that Hitler wanted to keep under direct control was useful in a general way for controlling the Atlantic coast. It also intersected with multiple north-south rail lines that were useful for controlling southern France. As a sovereign neutral, the planned French rump state would not be in a position to lend its transportation system to the Germans. Hence, German logistical needs required the direct control of these railways.

The documentary evidence presented here, together with railway maps from the period, permits us to describe the approximate path of the rail connection the Germans intended to keep in their zone of occupation (refer to map 1).⁸⁴ The north-south route described in Hitler’s June 17 meeting is the Paris-Bordeaux line, which departed from Paris, ran through Tours, Poitiers, Angoulême, and Libourne, to Bordeaux, whence it continued south to the Spanish border. In the departments of Charente and Vienne, the LoD ran approximately 20 kilometers east of this railway.

The east–west route described in the June 17 meeting was, at the time, the southernmost double-track railroad that cut across France north of the Mediterranean coast (refer to map 1). It began in Nantes, near the Atlantic coast, intersected the Paris-to-Bordeaux railroad at Tours, then zigzagged East through Bourges, Moulins, and Paray-le-Monial. East of Chagny, in eastern France, the French rail network became denser, with several double-track routes to the Northeast and Germany inter-connecting at multiple points, generating several possible itineraries for German logistical needs that satisfied Hitler’s requirement that they run entirely within the occupied zone. For the vast majority of its East–West path across central France, the LoD ran just to the south of this railway, normally within five kilometers.

This documentary and cartographic evidence strongly suggests that the Germans *did* have at least one explicit rationale for the local placement of the LoD: to keep important double-track railroads inside their directly occupied zone.⁸⁵ This strategic rationale for the placement of the LoD has serious consequences for any attempt to describe these data as the result of a natural experiment. If the LoD is not locally as-if random, then the communes on either side of it may have been unbalanced on a key confound: double-track railways the Germans deemed strategically significant and, therefore, the Resistance would have had good reason to attack. Were they?

Local Variation in Resistance: Strategic Logic versus Political Motives

To test whether the sample of communes near to the LoD was indeed balanced on this key confound, we return to

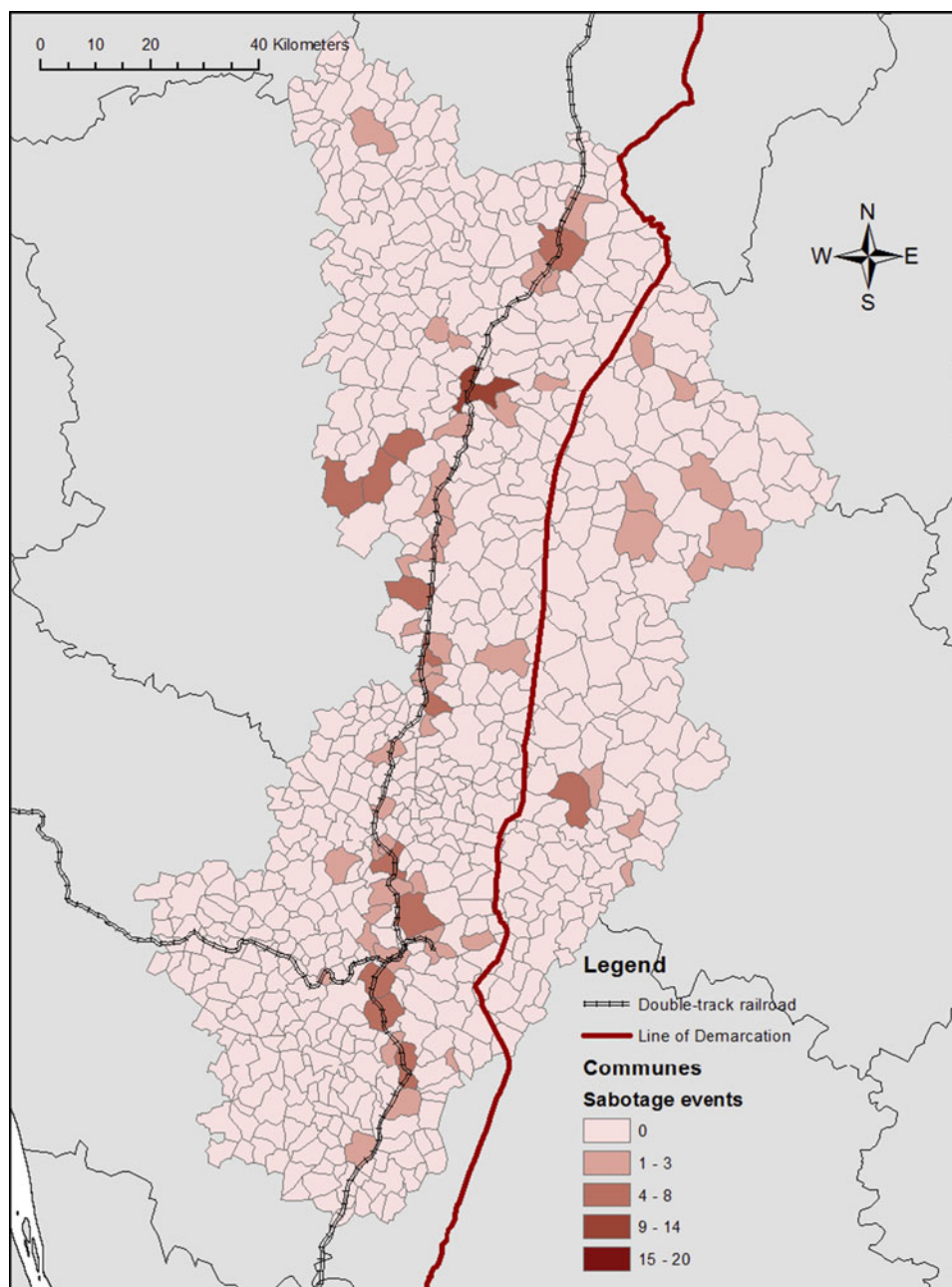
the data we introduced earlier. For each sabotage event reported in the original source, we attempted to identify the commune in which it occurred.⁸⁶

Maps 2–4 permit a closer look at all of the double-track lines and their proximity to the LoD in the departments of Charente, Cher, Saône-et-Loire, and Vienne, along with the distribution of sabotage. Three points emerge. First, sabotage was more common in the *zone nord* than in the *zone sud*, especially close to the LoD. Second, the pattern of attacks lies in spatial “strips” that closely follow the major railways. Elsewhere, sabotage was extremely rare. That sabotage occurred near railroads is unsurprising, given that 66 percent (378) of these events were described in the original sources as “railway sabotage.”⁸⁷ Third, we can see that major railways were *not* equally prevalent in the two zones. In the Charente and the Vienne, there were *no* major railways within the *zone sud*, while more than 200 kilometers of the Paris-Bordeaux railroad ran straight through the *zone nord*. In the Cher and the Saône-et-Loire, double-track railroads ran through both zones. But within tight bandwidths around the LoD, they were much more prevalent in the *zone nord*. Ironically, rather than increasing the homogeneity of the sample as RDDs are meant to do, the imposition of tight bandwidths around the LoD creates a biased selection of cases.⁸⁸

These maps bring together the key elements of our critique. If the French Resistance intentionally targeted double-track railways *and* the Germans intentionally sited the LoD to keep those same railways just inside the directly-occupied zone, then the positive statistical association between French communes’ rates of sabotage and their location in the *zone nord* is a classic instance of causal spuriousness. Both the supposed cause (*X*) and its effect (*Y*) are consequences of a third variable (*Z*). In this case, both the location of each commune in relation to the LoD (*X*) and the rate of sabotage attacks in each commune (*Y*) are the consequence the pre-treatment location of important railroads (*Z*).⁸⁹ We have shown persuasive evidence that the leaders of the Resistance in London created explicit plans to attack major railroads. We know that much of the Resistance had been organized under the command of the Free French leadership and that the plans to attack railroads had been transmitted to men on the ground in France, along with maps, weapons, explosives, and even Allied commandos to help them. We have good reason to think that crossing the LoD was a trivial matter at this time, such that motivated people near the line could easily attack targets on either side. We also have explicit documentary evidence that the Germans placed the LoD to keep important double-track railroads within the directly occupied zone, and we can show that the implementation of the LoD substantially conformed to those intentions. In short, we have identified a causal chain of interests, intentions, capabilities, and actions leading from a structural variable (railroad infrastructure) to a specific

Map 2

Double-track railways and sabotage in Charente and Vienne, 1940–1944



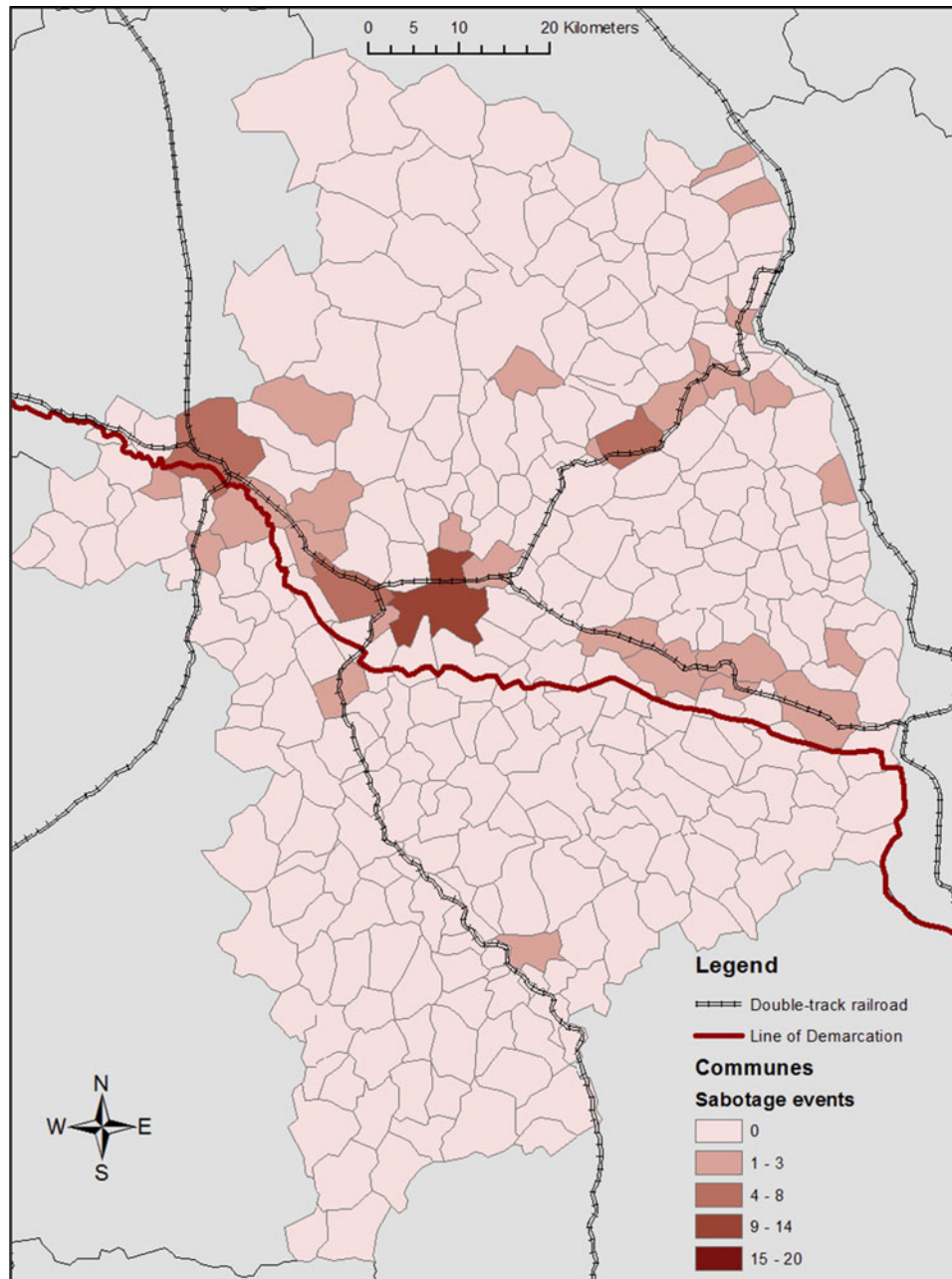
political behavior (violent resistance) that operates through a non-local process rather than through local motivations.

The causal chain we identify has several implications that we can test numerically, providing greater confidence in our claims. In table 3, we examine covariate balance on major railways. We code a binary variable—*Double track*—that equals one for communes intersected by a double-track railroad and zero otherwise. We then assume random

assignment of communes to the directly occupied and Vichy zones, and we conduct difference-of-means tests comparing *Double track* across these two conditions. In effect we ask: is assignment to either side of the LoD a good predictor of the location of double-track railways? We find that it is.

Table 3 shows two sets of comparisons. The top half drops all communes that intersect the LoD. This is a highly

Map 3
Double-track railways and sabotage in Cher, 1940–1944

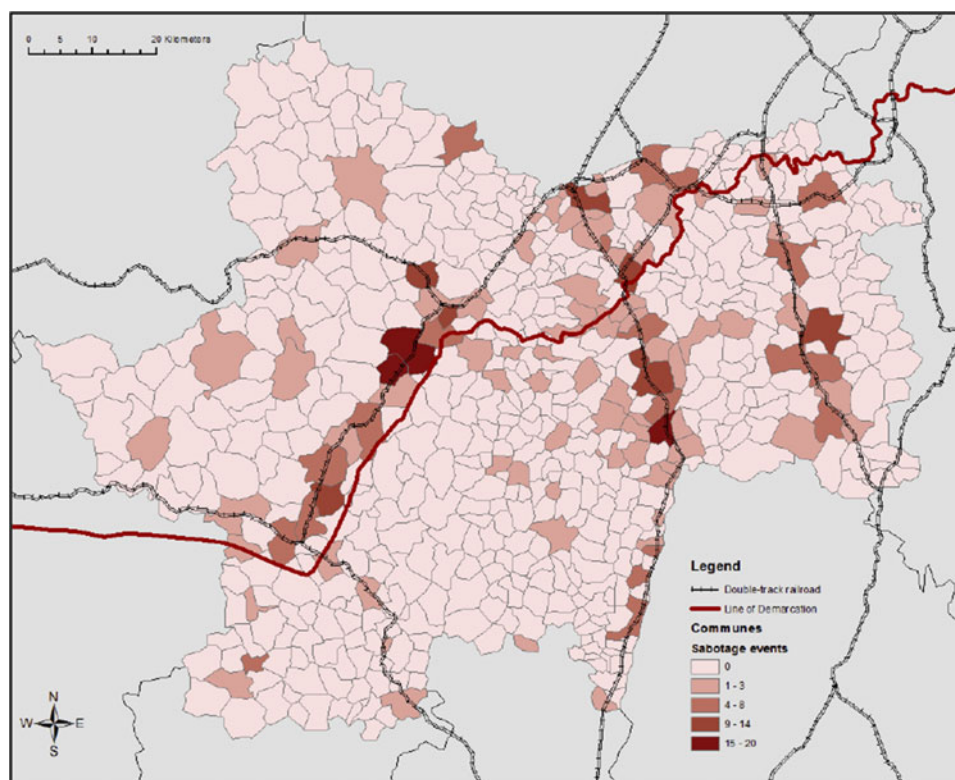


conservative assumption—i.e., it biases the analysis against finding an association between treatment assignment and the presence of double-track railroads—because for most of the communes intersected by the LoD *and* a double-track railroad, the railroad ran on the occupied side of the LoD.⁹⁰ For the four departments as a whole, and for 20- and 10-kilometer bandwidths around the LoD, treatment assignment (i.e., assignment to either side

of the LoD) is a statistically significant predictor of railroad location. The t-test at the 5-kilometer bandwidth is not significant, probably because so many of the communes in this tighter bandwidth were dropped from the analysis.⁹¹ Since we know on which side of the LoD the railroad ran for every commune intersected by the LoD, however, we can retrieve these observations. In the bottom half of table 3, we do so by splitting each of the communes

Map 4

Double-track railways and sabotage in Saône-et-Loire, 1940–1944



intersecting the LoD into an occupied and Vichy portion and coding *Double Track* for each of the resulting parts. When we do this, treatment assignment becomes a highly significant predictor of double-track railway location at the 5-kilometer bandwidth as well as at an even tighter bandwidth of 3 kilometers. Precisely where an RDD expects to maximize covariate balance—close to the discontinuity—we find major railways highly unbalanced.⁹²

In table 4 we examine the association between double-track railways and sabotage by taking the difference-of-means between communes intersected by double-track railways and communes without such railways, while controlling for occupation zones. All eight of these comparisons are substantively large and have the sign we expect. Across the various subsets of data created by conditioning on bandwidth and occupation zone, communes with double-track railroads had, on average, multiples of the number of attacks that occurred in communes without them. Of the eight comparisons, six are statistically significant at the 95 percent level. Within the Vichy zone, the comparisons at the 5- and 10-kilometer bandwidths are not significant at the 95 percent level. However, it is important to keep in mind that by excluding

communes intersected by the LoD, we omit some of the highest sabotage counts in the dataset. The significance of this point is indicated by the penultimate line in table 4, which makes the same comparison, but only for communes intersected by the LoD (the communes dropped in the previous eight comparisons). The difference of means is enormous at the LoD: communes with a double-track railroad had, on average, 33 times as many sabotage attacks as those without one. In the last line of table 4, we give the unconditional difference-of-means between communes with and without double-track railroads: the former had, on average, 16 times as many attacks as the latter.

Our numerical tests confirm both that major railways were systematically more common on the occupied side of the LoD than on the Vichy side and that communes with major railways on both side of the LoD were vastly more likely to experience sabotage than communes without such railways.

As a final numerical test of the pattern we identified in the maps, we compute differences-of-means in commune-level sabotage counts between the occupied and Vichy zones, while controlling for double-track railroads. Our results are reported in table 5.⁹³ We consider 5-, 10-, and 20-kilometer bandwidths around the LoD, as well as

Table 3**Difference of means test, commune intersects double-track railway conditional on treatment assignment**

		Mean of <i>Double track</i> , Occupied Zone (std. dev.)	Mean of <i>Double track</i> , Vichy Zone (std. dev.)	Difference of means, Occupied — Vichy	T-test (unequal variances)
Communes intersecting the Line of Demarcation dropped					
(1)	All Communes (N = 1418)	0.186 (0.390)	0.130 (0.336)	0.057	t = 2.922, p < 0.01
(2)	20Km Bandwidth (N = 691)	0.264 (0.442)	0.104 (0.305)	0.161	t = 5.610, p < 0.0001
(3)	10Km Bandwidth (N = 294)	0.277 (0.449)	0.144 (0.352)	0.133	t = 2.832, p < 0.01
(4)	5Km Bandwidth (N = 95)	0.327 (0.474)	0.239 (0.431)	0.087	t = 0.941, p = 0.349
Communes intersecting the Line of Demarcation split					
(5)	5Km Bandwidth (N = 353)	0.247 (0.433)	0.120 (0.326)	0.127	t = 3.123, p < 0.01
(6)	3Km Bandwidth (N = 253)	0.209 (0.408)	0.089 (0.285)	0.121	t = 2.731, p < 0.01

Two-tailed tests.

Table 4**Difference of means in *Sabotage* between communes with and without double-track railroads, conditional on zone of occupation**

		Mean of <i>Sabotage</i> , <i>Double track</i> = 0 (std. dev.)	Mean of <i>Sabotage</i> , <i>Double track</i> = 1 (std. dev.)	Difference of means, [<i>Double track</i> = 0] — [<i>Double track</i> = 1]	T-test (unequal variances)
Communes intersecting the LoD dropped					
5km Bandwidth	Occupied (N = 49)	0.030 (0.174)	3.125 (5.018)	−3.095	t = −2.466, p < 0.05
	Vichy (N = 46)	0.114 (0.323)	0.727 (1.489)	−0.613	t = −1.355, p = 0.204
10km Bandwidth	Occupied (N = 148)	0.084 (0.391)	2.122 (3.964)	−2.038	t = −3.286, p < 0.01
	Vichy (N = 146)	0.120 (0.502)	1.429 (3.234)	−1.309	t = −1.851, p = 0.079
20km Bandwidth	Occupied (N = 363)	0.082 (0.359)	2.063 (3.497)	−1.980	t = −5.538, p < 0.0001
	Vichy (N = 328)	0.129 (0.513)	2.235 (4.593)	−2.106	t = −2.672, p < 0.05
All Communes	Occupied (N = 832)	0.096 (0.506)	1.497 (2.944)	−1.401	t = −5.904, p < 0.0001
	Vichy (N = 586)	0.108 (0.454)	1.684 (3.480)	−1.576	t = −3.944, p < 0.001
Communes intersecting the LoD included					
Communes intersecting the Line only	(N = 130)	0.081 (0.340)	2.742 (4.195)	−2.661	t = −3.528, p < 0.01
All Communes	(N = 1548)	0.100 (0.474)	1.698 (3.282)	−1.599	t = −7.869, p < 0.0001

Two-tailed tests.

the entire departments. In general, the differences are small; in no case do they achieve statistical significance at the 95 percent level. For six of the eight comparisons, higher sabotage counts were observed in Vichy.⁹⁴ There is no evidence that a commune's assignment to either side of the LoD affects the incidence of sabotage once one controls for the presence of double-track railroads.^{95,96}

In sum, by delineating the LoD so that major railways were just inside the occupied zone, the Germans put more targets for the Resistance in the territory they occupied directly, but only close to the LoD. By focusing their analysis just on this small region and not taking

account of the strategic process through which communes were selected into the two zones, FM concluded that differences in the occupation regimes between the two sides of the LoD played a causally important role in determining the geographical pattern of violence. Our results indicate otherwise.

Conclusion: Causal Inference and Historical Research

If DBI research is going to usher in a “credibility revolution” in political methodology beyond the narrow set of questions amenable to experimental manipulation,

Table 5

Difference of means in *Sabotage* events between occupied and Vichy zones, conditional on intersection with a double-track railroad, with communes intersecting the LoD dropped

		Mean of <i>Sabotage</i> , Occupied Zone (std. dev.)	Mean of <i>Sabotage</i> , Vichy Zone (std. dev.)	Difference of means, Occupied– Vichy	T-test (unequal variances)
5km Bandwidth	<i>Doubletrack</i> = 1 (N = 27)	3.125 (5.018)	0.727 (1.489)	2.398	t = 1.799 p = 0.088
	<i>Doubletrack</i> = 0 (N = 68)	0.030 (0.174)	0.114 (0.323)	–0.084	t = –1.346 p = 0.184
10km Bandwidth	<i>Doubletrack</i> = 1 (N = 62)	2.122 (3.964)	1.429 (3.234)	0.693	t = 0.739 p = 0.464
	<i>Doubletrack</i> = 0 (N = 232)	0.084 (0.391)	0.120 (0.502)	–0.036	t = –0.611 p = 0.542
20km Bandwidth	<i>Doubletrack</i> = 1 (N = 130)	2.063 (3.497)	2.235 (4.593)	–0.173	t = –0.200 p = 0.843
	<i>Doubletrack</i> = 0 (N = 561)	0.082 (0.359)	0.129 (0.513)	–0.047	t = –1.263 p = 0.207
	<i>Doubletrack</i> = 1 (N = 231)	1.497 (2.944)	1.684 (3.480)	–0.187	t = –0.404 p = 0.687
All Communes	<i>Doubletrack</i> = 0 (N=1187)	0.096 (0.506)	0.108 (0.454)	–0.012	t = –0.423 p = 0.672

Two-tailed tests.

then careful historical scholarship must play a crucial role.⁹⁷ As we have shown, the key assumptions on which natural experiments depend are *causal* claims regarding unique events. As such, they can be validated only through idiographic research. If we were to believe, as many political scientists appear to do, that it is impossible to establish causation for unique historical events, then most natural experiments would lie on the wrong side of the demarcation line between scientific and non-scientific research. If, instead, we accept that idiographic causal inference is possible, then natural experiments are feasible, but their claim to a special epistemic status, clearly superior to conventional historical scholarship, is unsustainable.

Our examination of FM's study strongly supports the second perspective. Using old-fashioned historical research and simple statistical tests without an explicit identification strategy, we have provided a highly credible causal account of resistance action in the four departments in FM's study. We can say with great confidence why some communes were struck while others were spared; we are able to explain the timing of the majority of sabotage events; and we can account for the placement of the LoD. While we cannot precisely quantify the treatment effect of French communes' exposure to double-track railroads, it is not clear what such a quantity would be useful for. Almost certainly, the magnitude of this "local treatment effect" would vary from place to place within France and across other cases. Where and when railroads have played less of a vital role in military strategy, it is unlikely that we would observe similar patterns.

Our account of the causal processes that generated FM's data also permit us to explain why they arrived at unfounded conclusions. Beginning with validation of the as-if random character of the treatment assignment process, FM advance three claims in lieu of direct historical evidence: (i) existing work on the LoD does not explain its local placement;⁹⁸ (ii) the LoD cut across pre-existing boundaries;⁹⁹ and (iii) balance statistics on observed pretreatment covariates indicate no significant differences.¹⁰⁰ In the absence of evidence against the as-if random assumption (claims i and ii), FM rely on balance tests to support it (claim iii). But, as we have noted, balance on observed covariates cannot guarantee balance on variables that have not been observed. Attention to German intentions and actions reveals an explicit concern with control over major railways. Formal tests of the distribution of double-track railroads on opposite sides of the LoD then disclosed the highly unbalanced nature of FM's sample.

Second, FM's work illustrates the danger of reductionism in micro-level DBI research when validating the causal mechanism. By focusing on the micro-level dynamics at play in the small fragments of territory they study, FM missed the role played by the larger strategic context in producing violent resistance action. But incorporating broader levels of analysis demonstrates that sabotage events resulted from a concerted effort to liberate France by targeting strategic assets for the German war effort, namely double-track railroads. Social processes do not always play out at the level of the smallest spatial or political units we

can study. When the causal process that produce the data operates trans-locally, associations among political, economic, or social variables that are measureable on the disaggregated units may entirely miss the true underlying causes of the outcome under study. Researchers working on disaggregated data need to keep in mind the very real possibility that the causes of local phenomena are not themselves local, but rather may be determined far away.

Finally, FM's study exemplifies the role of qualitative evidence in validating the process through which the treatment is applied. The occupied and unoccupied portions of France were subject to distinctive governing institutions for nearly two and a half years, but our analysis of the timeline of sabotage shows that 95 percent of the attacks occurred after the German occupation of the *zone sud*, when the governing regimes on opposite sides of the LoD had converged. While there is some evidence of continuing differences between the zones, the precise nature of these differences is difficult to specify and does not appear to conform to a theoretical distinction between direct and indirect rule. The importance of process-tracing the causal story leading from the independent to the dependent variable in DBI research increases with the complexity of the causal process that supposedly links the two variables. When the causal process is very simple or direct, in the sense that it happens over a short time-span and through very short chains of agency, there is less need for in-depth process-tracing. When, on the contrary, the causal process plays out over years and requires deep, multi-actor and multi-level chains of agency, then threats to inference multiply and extensive process-tracing becomes indispensable to validate the treatment application.

The general lesson for DBI to be drawn from our discussion of FM's analysis could not be clearer: there is no substitute for reconstructing the data-generating process through careful historical research. Consequently, we need to strengthen the scholarly criteria for validating DBI research. Here, we call for a more historically-informed form of DBI that acknowledges the primary role of qualitative evidence in producing scientific knowledge of political phenomena.

As a practical matter, most natural experiments are conducted by non-historians who lack deep knowledge of the countries and time periods under investigation. Likewise, the journal editors and referees who vet their work are typically experts in the relevant theories and methods, not the historical or sociological context in which a natural experiment is embedded. Even when such expertise is present, the relevant data may be so specialized that only a researcher determined to evaluate a given piece of DBI scholarship will be able to detect important mistakes.

To overcome these limitations, we call for greater interaction between researchers with deep historical or contextual knowledge and those with the high-tech skills

necessary to conduct DBI research, taking seriously the spirit of analytic eclecticism and methodological pluralism (including both qualitative and quantitative approaches) when evaluating DBI research. The burden of proof in validating DBI research falls above all on the researcher, who should always provide qualitative evidence of the process of treatment assignment and administration—in the form of treatment-assignment, independent-variable, and mechanism CPOs. But others have a part to play as well. Journal editors and referees should require that researchers document, as directly as possible, the most relevant causal processes. They should refuse to accept the absence of evidence for a strategic rationale as evidence of its absence unless a great many stones have been overturned. Above all, they should require that this information be packaged in a manner accessible to researchers with specialized contextual knowledge of each particular case—and journal editors should have DBI research vetted by referees deeply knowledgeable on the historical and sociological processes at work.

Given the highly-specialized historical and sociological knowledge required to validate much DBI scholarship, however, it may not be realistic to expect that journal referees and editors will catch even fairly significant errors. Establishing that a journal article's quantitative results can be reproduced from the authors' data and software code is a useful, but insufficient, standard of replication. For DBI research in particular, the real action often lies elsewhere: in archival materials, elite interviews, and foreign-language secondary sources. Meaningful post-publication review of DBI research will often require substantial original research, which current disciplinary incentives do not reward. To change this unfortunate state of affairs, the journals with the highest status in the profession should adopt policies encouraging the publication of forums on, and rejoinders to, previously published articles. The absence of such policies postpones the credibility revolution.

The new prominence of DBI research offers a choice that is in many ways similar to several others that political scientists have faced in the past. In response to successive attempts at establishing a line of demarcation in the discipline privileging particular approaches—quantitative observational studies in the 1960s, game-theoretic models in the 1990s—political science has remained faithful to the methodological pluralism that is today one of the hallmarks of the discipline. Given the extremely narrow range of political issues that can be studied using RCTs, we need once again to reaffirm the central role of pluralism in the study of politics. We hope that establishing the epistemic parity between historical and DBI-research—respectively, the least and most scientific observational methodologies according to proponents of the DBI revolution—will contribute to ensuring that pluralism remains a core value in political science.

Notes

- 1 Przeworski 2007, 168.
- 2 One role on which DBI advocates attribute great utility to qualitative or historical research is in the validation of natural experiments. See Dunning 2012, 2015. We discuss this later.
- 3 See Gerber and Green 2012, 5–7, 37–39; Dunning 2012, 5–6; Sekhon 2009; Samii 2016.
- 4 See King, Keohane, and Verba 1994.
- 5 See Freedman 1991.
- 6 See Dunning 2014, 231.
- 7 See Ferwerda and Miller 2014.
- 8 *Ibid.*, 658.
- 9 See Dunning 2014, 227.
- 10 See Dunning 2012, 212–218.
- 11 Dunning 2014, 212. Following Collier, Brady, and Seawright (2004, 252–258) we understand CPOs to be facts established through observation that give us inferential leverage in establishing causation. Causation, as such, is never directly observed.
- 12 See Dunning 2012, 239–243.
- 13 Hyde 2007, 46. See also Dunning 2012, 28.
- 14 *Ibid.*, 13.
- 15 *Ibid.*, 28.
- 16 Dunning 2014, 212.
- 17 Gerber and Green 2012, 5.
- 18 Imbalance on observed variables is, however, important evidence *against* a natural experiment; see Caghey and Sekhon 2011.
- 19 See Mahoney 2010; Dunning 2012, chapter 7.
- 20 *Ibid.*, 209.
- 21 Mahoney 2010, 125.
- 22 Dunning 2012, 209, emphasis original.
- 23 Collier, Brady, and Seawright 2004, 254–255.
- 24 Ferwerda and Miller 2014.
- 25 This sub-section builds on Kocher, Lawrence, and Monteiro 2013.
- 26 See Paxton 2001, 279; Jackson 2001, 414.
- 27 See Kedward 1985, 56–57.
- 28 See Gildea 2002, 323–324.
- 29 See Kedward 1985, 49–50.
- 30 See Michel 1950, 23. Wright 1962, 340–341; Jackson 2001, 417; and Kedward 1978, 137 make the related but distinct claim that the Resistance in the unoccupied zone had to be framed in opposition to the collaboration of the regime, while in the occupied zone, where the regime's reach was limited, Vichy was not as relevant.
- 31 The sources we consulted include 86 of the 90 departments of metropolitan France, excluding only the departments of Seine, Seine-et-Oise, the island of Corsica, and the Bas-Rhin, which was annexed to Germany in 1940.
- 32 See Lachapelle 1936, viii, 328–332.
- 33 These results are consistent with the argument advanced in Kocher, Lawrence, and Monteiro 2013.
- 34 Ferwerda and Miller 2014, 658.
- 35 See Dunning 2014.
- 36 See Général de la Barre de Nanteuil 1973–1986. Ferwerda and Miller 2014, 648, analyze two types of events, *sabotage* and *fighting* (primarily attacks by the Resistance on German forces or French collaborators). They report statistically and substantively significant results for *sabotage*, but much weaker results for *fighting*; hence, we focus on *sabotage*. We recovered complete dates for 686 out of 732 total sabotage events.
- 37 These percentages exclude 46 events for which we do not have complete date information. Of these, at least 35 happened in 1944.
- 38 From November 1942 until the end of the war, we follow the canonical usage in referring to the part of France occupied from 1940 as the “zone nord” and the formerly unoccupied zone as the “zone sud.”
- 39 Jackson 2001, 228.
- 40 See Alary 2003, 269–273.
- 41 Gildea 2015, 285.
- 42 André 2013, 14–16.
- 43 See Jäckel 1968, 366–368.
- 44 See *Ibid.*, 366; Kedward 1993, 14.
- 45 Jäckel 1968, 368.
- 46 Ferwerda and Miller 2014, 646.
- 47 Paxton 2001(1972), 281.
- 48 Jäckel 1968, 371.
- 49 Jackson 2001, 232.
- 50 Gildea 2002, 51–56.
- 51 In reality, these were not discrete policies, but rather amalgams of many interacting policies, of which native authority was only one. When treatments are “compound” in this sense, causal identification depends on the assumption that only the treatment of interest affects potential outcomes (Keele and Tiunuk 2014, 7–10, n. 8). This requirement is analogous to the exclusion restriction in the context of instrumental variables.
- 52 Gildea 2002, 417.
- 53 Jackson 2001, 554.
- 54 See Veyret 2001, 49–51.
- 55 See *Ibid.*, 51.
- 56 See Calmon 2000, 53; Durand 1968, 426–427; Veyret 2001, 129–132.
- 57 Durand 1968, 428.
- 58 *Ibid.*, 427.
- 59 Calmon 2000, 50.
- 60 *Ibid.*, 52.
- 61 *Ibid.*, 50–52.
- 62 *Ibid.*, 52–53.
- 63 *Ibid.*, 53–54.
- 64 See Veyret 2001, 49–50.

- 65 See Ibid., 82–83.
- 66 See Ibid., 120.
- 67 See Ibid., 110, 120.
- 68 Ibid., 111–112.
- 69 Quoted in Ibid., 132.
- 70 Ferwerda and Miller 2014, 642.
- 71 Ibid., 647.
- 72 Ibid., 645.
- 73 Ibid., 645, fn. 9.
- 74 Jodl was Chief of the Operations Staff of the Armed Forces High Command (*Oberkommando der Wehrmacht*, or OKW). Keitel was a member of the German delegation to the Armistice negotiations and signed the instrument for Germany.
- 75 Böhme 1966, 21, our translation. The author of the report is General Hermann Böhme, who was, in June 1940, a top staff officer in the OKW bureau that did the preparatory work for this meeting. Planning for the LoD—which mapped France “perfectly in its smallest details with its police stations, its industrial bakeries, its steel mills, etc.”—began prior to the invasion; Alary 2007, 7.
- 76 See Böhme 1966, 20–21.
- 77 Germany, Auswärtiges Amt 1961, 334. This quotation is from a French translation of the original German report on the meeting prepared for the German foreign ministry.
- 78 See Ibid., 305.
- 79 American Journal of International Law 1940, 173.
- 80 See *New York Times*, June 26, 1940; *Le Matin*, June 27, 1940.
- 81 A French version of this note was published by *Le Matin* on June 27, 1940. Multiple transcriptions of the Armistice available online include this note as an Appendix. Available at: <http://avalon.law.yale.edu/wwii/frgearm.asp>. Last accessed May 5, 2016. See Texte de l’armistice signé à Rethondes le 22 Juin, 1940. Available at: <http://mjp.univ-perp.fr/france/1940armistice.htm>. Last accessed May 5, 2016.
- 82 See van Creveld 1977, 142–147; Overly 1973; Veyret 2001.
- 83 See van Creveld 1977, 158.
- 84 Our maps and the geographical data analyzed in the next section derive from two principal sources: the GADM database of Global Administrative Areas (<http://gadm.org/home>), and a 1:200,000-scale map series depicting the LoD that we obtained from the Cartothèque of the Institut Géographique Nationale (IGN) of France.
- 85 Just how local German decision-making would have to have been to affect FM’s results is an empirical question that we tackle later. It is worth noting, however, that in many parts of the Cher and the Saône-et-Loire, shifting the LoD less than 10 kilometers north would have left the railroad that accounts for most of the sabotage attacks in the opposite zone. In the Charente and the Vienne, moving the LoD 20 kilometers west would have placed it just about on top of the Paris-Bordeaux railway. The definition of “local” implied by FM’s 5–25 kilometer bandwidths suggests that these decisions could be important.
- 86 We were able to assign communes to 573 *Sabotage* events and 613 *Fighting* events, while FM identified 598 *Sabotage* events and 595 *Fighting* events. Roughly 7 percent of the *Sabotage* events and 10 percent of the *Fighting* events could not be assigned to a unique commune. As in the section on the logic of violent resistance, our analysis focuses on *Sabotage*. Refer to Appendix A, tables A.2 and A.3 for numerical tests on *Fighting*. FM’s replication data became available after we completed our analysis. We have since examined the two datasets side-by-side. Refer to Appendix B for a discussion of some discrepancies as well as a replication of our key results using their data. All our key criticisms of their analysis are supported.
- 87 Information on the type of sabotage is missing for an additional 23 percent (130) of these events, leaving open the possibility that as many as 89 percent were directed at railroads.
- 88 FM compute balance statistics and match on a variable they label *Train distance*, which measures each commune’s distance to the nearest train station. The French railway network of the 1940s was extensive, however, and it included many strategically irrelevant single-track and narrow-gauge lines. Thus, *Train distance* does not capture the strategically significant double-track railways that directly influenced German plans for the LoD.
- 89 In somewhat different language, if our account is correct then FM’s estimate of the local average treatment effect (LATE) is biased due to unobserved heterogeneity. Formally, let Y be the number of resistance events, let X be the indicator of assignment to the occupied zone or the Vichy zone, and let i and j indicate, respectively, untreated and treated units. FM observe $Y_i|X_i = 0$ and $Y_j|X_j = 1$. To recover the LATE, the key assumption is that $E(Y_i|X_i = 0) = E(Y_j|X_j = 0)$. That is, in the absence of treatment, the treated units would have had the same incidence of violence as the units that were in fact treated.
- 90 Twenty-four percent (31/130) of communes intersected by the LoD were also intersected by a double-track railroad. Of those, the railroad ran only on the occupied side in 68 percent (21/31).
- 91 As robustness tests, we repeated this analysis using the length (in kilometers) of double-track railroad and the shortest distance from each commune’s

boundary to the closest double-track railroad (refer to Appendix A, tables A.4 and A.5, respectively), in each case with similar results. Using entirely different British maps, Ferwerda and Miller (2015, 26, table A.1) also find multiple-track railroads sharply unbalanced across the LoD.

- 92 If railroads are unbalanced, one might wonder if other covariates are similarly unbalanced. Railway lines are not built just anywhere. They often connect existing cities and towns. They influence economic and population growth as people and industries move to locations where transportation is convenient. They require rights-of-way, which tend to accrue additional infrastructure: fuel and water pipelines, electricity and telephone wires. Natural geography makes some routes more feasible and less costly than others. Civil engineers site them next to water courses and along the contours of the land to avoid building tunnels and bridges. We examined some readily-available covariates and found that population density is unbalanced across the two zones for every bandwidth wider than 10 km; population density is also highly associated with both railroad location and the rate of sabotage.
- 93 Refer to Appendix A for (i) our formal estimand, (ii) our defense of the difference-of-means estimator instead of local linear regression, and (iii) local linear regression estimates as a robustness check.
- 94 While statistically insignificant, the inter-zone difference among communes with double-track railroads for the 5-km bandwidth is substantively quite large (see the boxed cell in table 5). This comparison zeroes in on 27 communes right along the LoD in the departments of Cher and Saône-et-Loire. Much of this difference is driven by two communes (Montceau-les-Mines and Montchanin) that together account for 30 of the 58 sabotage attacks in these 27 communes. The low statistical power of this comparison stems mainly from the high variance created by these two outliers, which were extremely unusual among communes in western Burgundy in a number of respects (size, population density, industrialization, labor mobilization) that are very likely to have been causally related to railroad infrastructure.
- 95 For an analysis similar to the one conducted above in tables 4–5, but restricted to the pre-D-Day period, refer to tables A.6–A.7 in Appendix A. There are no important differences in the results.
- 96 Ferwerda and Miller (2015, 26, table A.1) find no statistically significant differences between the two zones in the rate of sabotage per kilometer of multiple-track railroad, confirming our results.
- 97 See Angrist and Pischke 2010.
- 98 See Ferwerda and Miller 2014, 645, n. 9.

99 Ibid., 645.

100 Ibid., 647–648.

Supplementary Materials

Online Appendix A

- Descriptive statistics on determinants of sabotage in wartime France
- Analysis of fighting
- Additional tests on communes and double-track railways
- Analysis of sabotage restricted to pre-D-Day period
- Discussion of FM's use of local linear regression as their estimator

Online Appendix B

- Discussion of differences between our data and FM's replication data

Online Appendix C

- Response to Ferwerda and Miller (2015)

Replication data and code

<http://dx.doi.org/10.1017/S1537592716002863>

References

- Alary, Eric. 2003. *La Ligne de Demarcation (1940–1944)*. Paris: Perrin.
- . 2007. “La Ligne de Démarcation.” *Revue d'Histoire du Pays Chatelleraudais* 13: 5–23.
- American Society of International Law. 1940. “Armistice between France and Germany.” *American Journal of International Law* (Supplement: Official Documents) 34(4): 173–78.
- André, Philippe. 2013. *La Résistance confisquée? Les délégués militaires du général de Gaulle de Londres à la Libération*. Paris: Éditions Perrin.
- Angrist, Joshua D. and Jorn-Steffen Pischke. 2010. “The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics.” *Journal of Economic Perspectives* 24(2): 3–30.
- Böhme, Hermann. 1966. *Der deutschfranzösische Waffenstillstand im zweiten Weltkrieg*. Stuttgart: Deutsche Verlagsanstalt.
- Calmon, Jean-Henri. 2000. *Occupation, Résistance, et Libération dans la Vienne en 30 Questions*. La Crèche: Geste Éditions.
- Caughey, Devin and Jasjeet Sekhon. 2011. “Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008.” *Political Analysis* 19(4): 385–408.
- Collier, David, Henry E. Brady, and Jason Seawright. 2004. “Sources of Leverage in Causal Inference.” In *Rethinking Social Inquiry: Diverse Tools, Shared*

- Standards*, ed. H. Brady and D. Collier. Rowman & Littlefield.
- Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.
- . 2014. "Improving Process Tracing: The Case of Multi-Method Research." In *Process Tracing: From Metaphor to Analytic Tool*, ed. Andrew Bennett and Jeffrey Checkel. Cambridge: Cambridge University Press.
- Durand, Paul. 1968. *La S.N.C.F. pendant la guerre: sa résistance à l'occupant*. Paris: Presses Universitaires de France.
- Ferwerda, Jeremy and Nicholas L. Miller. 2014. "Political Devolution and Resistance to Foreign Rule: A Natural Experiment." *American Political Science Review* 108(3): 642–60.
- . 2015. "Rail Lines and Demarcation Lines." Working paper.
- Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology* 21: 291–313.
- Général de la Barre de Nanteuil, ed. 1973–1986. *Historique des Unités Combattantes de la Résistance 1940–1944*. Vols. 1, 9, 11, 12, 18, 20. Vincennes: Ministère des Armées, Etat-major de l'Armée de Terre, Service Historique.
- Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton & Co.
- Germany, Auswärtiges Amt. 1961. *Archives secretes de la Wilhelmstrasse*. Vol. 9, pt. 2. Trans. Michel Tournier. Paris: Éditions Plon.
- Gildea, Robert. 2002. *Marianne in Chains: In Search of the German Occupation, 1940–1945*. London: Macmillan.
- . 2015. *Fighters in the Shadows: A New History of the French Resistance*. Cambridge, MA: Harvard University Press.
- Hyde, Susan D. 2007. "The Observer Effect in International Politics: Evidence from a Natural Experiment." *World Politics* 60(1): 37–63.
- Jäckel, Eberhard. 1968. *La France dans l'Europe de Hitler*. Trans. Denise Meunier. Paris: Librairie Arthème Fayard.
- Jackson, Julian. 2001. *France, The Dark Years, 1940–1944*. New York: Oxford University Press.
- Kedward, H. R. 1978. *Resistance in Vichy France: A Study of Ideas and Motivation in the Southern Zone 1940–1942*. Oxford: Oxford University Press.
- . 1985. *Occupied France: Collaboration and Resistance 1940–1944*. Oxford: Basil Blackwell, Ltd.
- . 1993. *In Search of the Maquis: Rural Resistance in Southern France, 1942–1944*. Oxford: Clarendon Press.
- Keele, Luke and Rocio Titunuk. 2014. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23(1): 127–55.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Kocher, Matthew Adam, Adria K. Lawrence, and Nuno P. Monteiro. 2013. "The Rabbit in the Hat: Nationalism and Resistance to Foreign Occupation." Working paper, Yale University.
- Lachapelle, Georges. 1936. *Elections legislatives, 26 avril & mai, 1936: résultats officiels*. Paris: Le Temps.
- Mahoney, James. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62(1): 120–47.
- Michel, Henri. 1950. *Histoire de la Résistance en France (1940–1944)*. Paris: Presses Universitaires de France.
- Overy, Richard J. 1973. "Transportation and Rearmament in the Third Reich." *Historical Journal* 16(2): 389–409.
- Paxton, Robert O. 2001[1972]. *Vichy France: Old Guard and New Order, 1940–1944*. New ed. New York: Columbia University Press.
- Przeworski, Adam. 2007. "Is the Science of Comparative Politics Possible?" In *The Oxford Handbook of Comparative Politics*, ed. Carles Boix and Susan Stokes. Oxford: Oxford University Press.
- République Française, Direction de la Statistique Générale. 1938. *Résultats statistique du recensement générale de la population effectué le 8 mars 1936*. 3 vols. Paris: Imprimerie Nationale.
- Samii, Cyrus. 2016. Causal Empiricism in Quantitative Research. *Journal of Politics* 78(3): 941–55.
- Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." *Annual Review of Political Science* 12: 487–508.
- Van Creveld, Martin. 1977. *Supplying War: Logistics from Wallenstein to Patton*. Cambridge: Cambridge University Press.
- Veyret, Patrick. 2001. *Mémoire de la Résistance en Saône-et-Loire: Maquis, Forces Spéciales, et SAS*. Châtillon-sur-Chalaronne: Editions La Taillanderie.
- Wright, Gordon. 1962. "Reflections on the French Resistance (1940–1944)." *Political Science Quarterly* 77(3): 336–49.