Check for updates

Original Article

Sociological Methods & Research
1–36
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00491241221082597
journals.sagepub.com/home/smr
SAGE

Do Quantitative and Qualitative Research Reflect two Distinct Cultures? An Empirical Analysis of 180 Articles Suggests "no"

David Kuehn [1]
and Ingo Rohlfing [2]

Abstract

The debate about the characteristics and advantages of quantitative and qualitative methods is decades old. In their seminal monograph, *A Tale of Two Cultures* (2012, *ATTC*), Gary Goertz and James Mahoney argue that methods and research design practices for causal inference can be distinguished as two cultures that systematically differ from each other along 25 specific characteristics. ATTC's stated goal is a description of empirical patterns in quantitative and qualitative research. Yet, it does not include a systematic empirical evaluation as to whether the 25 are relevant and valid descriptors of applied research. In this paper, we derive five observable implications from ATTC and test the implications against a stratified random sample of 90 qualitative and 90 quantitative articles published in six journals between 1990–2012. Our analysis provides little support for the two-cultures hypothesis. Quantitative methods are largely implemented as described in ATTC, whereas qualitative methods are much

[1]GIGA German Institute for Global and Area Studies, Hamburg, Germany
[2]Cologne Center for Comparative Politics, University of Cologne, Köln, Germany

Corresponding Author:
David Kuehn, GIGA Institute for Global and Area Studies, Hamburg, Germany.
Email: david.kuehn@giga-hamburg.de

more diverse than ATTC suggests. While some practices do indeed conform to the qualitative culture, many others are implemented in a manner that ATTC characterizes as constitutive of the quantitative culture. We find very little evidence for ATTC's anchoring of qualitative research with set-theoretic approaches to empirical social science research. The set-theoretic template only applies to a fraction of the qualitative research that we reviewed, with the majority of qualitative work incorporating different method choices.

Keywords

causal inference, causality, counterfactuals, descriptive statistics, hypothesis testing, observational studies

The debate about the relative pros and cons of quantitative and qualitative methods is decades old and continues to be lively (Gerring 2017; Mahoney 2010). Within this debate, the monograph *A Tale of Two Cultures (ATTC)* takes stock of and offers a novel perspective by mapping "typical practices" in the application of qualitative and quantitative methods in political science and sociology (Goertz and Mahoney 2012: 10). ATTC argues that qualitative and quantitative research practices can best be characterized as two *cultures* that systematically differ from each other. Henceforth, we refer to this as the *two-cultures hypothesis* (or argument). The cultures are argued to be based on different "values, beliefs, and norms" and are "associated with distinctive research procedures and practices" (Goertz and Mahoney 2012: 1). Research practices are broadly construed, with five dimensions of empirical research subsuming 25 individual items that capture a variety of decisions and choices researchers must make in their work.[1] The 25 items comprise fundamental beliefs about the nature of causality (symmetric vs asymmetric), the formation of concepts, various research design elements and practical decisions such as how to choose cases. For ease of discussion, we refer to them as matters of method implementation, application, or practice. The claim of fundamental differences between qualitative and quantitative methods is hardly novel in the social sciences (for example, Brady and Collier 2010). However, ATTC delivers by far the most articulate and comprehensive discussion, making one of the strongest claims about the scope and relevance of these differences.

In this paper, we perform a comprehensive empirical test of the two-cultures hypothesis with the goal of moving forward the debate about qualitative and quantitative methods practices. A test of the two-cultures hypothesis is important because there are three related reasons that ATTC influences empirical and methods-oriented research.[2] First, the two-cultures hypothesis is referenced in empirical and methods-related research as if it was an empirical fact rather than a hypothesis (Blatter and Haverland 2013; Koivu and Damman 2015; Schneider and Wagemann 2013).[3] Any hypothesis needs to be evaluated in light of empirical evidence before it is accepted as being probably correct.

Second and related, discussants of ATTC express doubts that it correctly describes methods applications (Brady 2013; Elman 2013; Rohlfing 2013). These doubts, however, have not been substantiated with systematic empirical evidence. Arguments for and against the existence of two cultures have thus far been confined to illustrative examples that are easy to find for both proponents and sceptics of the two-cultures argument.[4] We aim to overcome this impasse with a comprehensive empirical test.[5]

Third, ATTC is at times cited as if it was suggesting *best practices* in the application of methods, and is criticized for promoting incorrect best practices (Beach and Pedersen 2013b).[6] ATTC cannot be blamed for such misreadings because it makes very explicit that its main focus is *description*, not *prescription*. Nonetheless, an empirical analysis of how methods are applied and whether this conforms to the two-cultures hypothesis is important because it highlights what ATTC is about and might help to correct any misinterpretations. Our article does not aim to make prescriptive arguments about method application in political science, of which there already quite a number, but to get a better understanding about how it is done.[7]

The present analysis goes significantly beyond a previous empirical test of the two-cultures hypothesis by Kuehn and Rohlfing (2016), both conceptually and empirically. In conceptual terms, the 2016 study focuses on a single observable implication: namely whether research practices in published qualitative and quantitative articles conform to the descriptions in ATTC. In our article, we formally derive four additional implications from the two-cultures hypothesis and test them. Empirically, the earlier paper is based on just 30 articles, 15 quantitative and 15 qualitative, from three journals (*Comparative Political Studies, European Journal of Political Research, World Politics*), published between 2008 and 2012. Our article is based on a research sample of 180 articles plus a sensitivity test involving 30 additional qualitative articles from a variety of journals. In terms of methods, the earlier analysis only cross-tabulates observed methods practices across all 30 articles

against the type of method (qualitative or quantitative) that is applied in a study. The cross-tabulations cast doubt on the validity of the two-cultures hypothesis because the distribution of method practices deviates from the expectations. However, an analysis of the 30 articles can only be considered preliminary because of the small sample size, its limitation to a single observable implication, and the simple empirical approach.

We proceed as follows. After a brief summary of the two-cultures argument (section "The two-Cultures Hypothesis"), we derive five observable implications for which we should find empirical evidence if the hypothesis was correct (section "Five Observable Implications of the two-Cultures Hypothesis"). The empirical analysis is based on an in-depth content analysis of a random sample of 90 qualitative and 90 quantitative empirical articles that are stratified by three time periods and six major journals covering different subfields of political science (section "Empirical Strategy").[8] We coded each article independently of each other with respect to the 25 methods practices as defined in ATTC (see also Appendix C). The codings constitute the data that we use to assess the five observable implications in sections "Observable Implication 1: Relevance of Method Practices" to "Observable Implication 5: Method Applications Over Time".

We find that none of the five observable implications is fully empirically supported. Quantitative methods practice largely meets ATTC's characterization of the quantitative culture. We find a number of exceptions, such as the modeling of interactions in a sizeable share of quantitative articles, which is theorized to be a feature of qualitative research. Qualitative research is more diverse than expected and often does not display the characteristics of the qualitative culture. While some practices do indeed conform to the qualitative culture (e.g., the number of cases studied), many others are implemented in a manner that ATTC characterizes as constitutive of the quantitative culture.

We believe that the disagreement between our empirical findings and the expectations formulated in ATTC is due to its fundamental assumption that qualitative research is anchored in a set-theoretic framework. This assumption is reflected, for example, in how ATTC discusses the asymmetric conceptualizations of variables; causal relations in terms of necessity and sufficiency; and the conjunction of conditions that produce the outcome. Moreover, among the 25 indicators for the qualitative culture, ATTC includes several methods practices that are relevant to one specific subset of set theoretic methods, namely Qualitative Comparative Analysis (QCA), but are not relevant for research practices in non-QCA qualitative research. These include inter alia the organization of data in truth tables, and the interpretation of

triangular data as evidence of set relations.[9] In total, nine of the 25 qualitative research practices are based on the idea of set-theoretic research.

Section "Summary and Alternative Explanations" summarizes the findings and discusses potential alternative explanations for those findings other than the two-cultures hypothesis being incorrect. We present evidence indicating that the findings are neither the result of a particular bias in the selected journals, nor that the finding is confined to the periods under review. We also address in detail why potentially implicit methods practices, especially in qualitative research, are unlikely to affect our results. The final section concludes by summarizing an exploratory empirical cluster analysis of the 90 qualitative articles. We find that empirical research in the qualitative tradition is considerably more diverse than the two-cultures hypothesis suggests. There is evidence for the existence of at least three "sub-types" of qualitative methods applications. Articles that are based on set-theoretic assumptions constitute the numerically smallest cluster.

The Two-Cultures Hypothesis

The core argument of ATTC is that the application of methods in the social sciences systematically differs between qualitative and quantitative empirical research. The claim of systematic differences has been made many times before (for example, Beach and Pedersen 2013a, chapter 2; Collier et al. 2004; George and Bennett 2005, chapter 1). ATTC goes significantly beyond previous work and contends that the two methods follow different, coherent practices that can be located on five dimensions, capturing a total of 25 individual design and method decisions (see Appendix A).

In a stylized perspective, the argument is that qualitative research is about explaining the outcome of individual cases; follows a set-relational understanding of causality; tends to engage in in-depth analysis of a small number of cases with the aim of generalizing findings to a narrowly defined context; understands the measurement of the operational variables as a semantic process of defining and bringing into logical relationships a multitude of concept dimensions; and develops asymmetrical causal arguments. In contrast, quantitative research does not primarily aim to explain the outcome of individual cases; understands causality in terms of the average treatment effect of individual variables and, only rarely, interaction terms; quantitative studies usually cover a large number of cases to allow for broad generalizations; conceives of variables as latent variables that are operationalized with indicators; and develops "symmetric causal arguments

in which the same variables and model explain the presence versus absence of an outcome" (Goertz and Mahoney 2012: 225).

ATTC is exemplarily clear in describing the two cultures and defining their dimensions and constitutive attributes. At the same time, the discussion falls short of demonstrating that the two cultures are empirically valid descriptions of qualitative and quantitative research. The book provides well-chosen discussions of selected methods applications and examples that fully illustrate the two-cultures argument. However, these illustrations do not constitute the kind of empirical evidence that is necessary to making a convincing case for the empirical correctness of the two-cultures hypothesis.

The appendix of ATTC provides a survey of 216 articles from the *American Journal of Sociology, American Political Science Review, American Sociological Review, Comparative Politics, International Organization,* and *World Politics*. Substantively, it identifies whether an article uses predominately qualitative or quantitative methods and what type of method in particular (case studies, QCA, OLS regression, etc.). This survey is an insightful first step, but does not constitute a systematic evaluation of the much more nuanced thesis of two quantitative and qualitative cultures that differ from each other with regard to the 25 method practices.

Five Observable Implications of the two-Cultures Hypothesis

The two-cultures hypothesis is a statement about common method practices in the social sciences that can be tested empirically. However, the main hypothesis is too general and broad for a direct empirical test. In this section, we use it as the basis for deriving five observable implications for which we should find empirical evidence if the two-cultures hypothesis was correct.

First, we should observe that the 25 method practices constituting the two cultures are conceptually valid. A methods practice is conceptually valid for the two-cultures hypothesis when it is frequently or, ideally, always observed in quantitative and qualitative research. What is barely relevant in the practice of empirical research cannot be relevant for distinguishing cultures and is evidence for the inclusion of irrelevant practices into their definitions. The finding that one or multiple practices are rare in empirical work would not disconfirm that qualitative and quantitative methods differ on the subset of practices that are regularly exercised. However, it would suggest that the two

methods cultures are not constituted by all 25 practices defined in ATTC. This first implication is the most general one because we only ask *whether* the design and method decisions are empirically relevant as a necessary prerequisite for asking *how* they are implemented. The how-question is asked in different ways to motivate the other four implications.

Second, if the two-cultures hypothesis was true, it implies that "methods culture" is a single *latent* dimension underlying the use of methods in empirical qualitative and quantitative research. A confirmatory factor analysis should show that the 25 methods practices can be reduced to a single latent methods dimension.

Third, if the 25 methods practices were to differ systematically *between* and were uniformly implemented *within* quantitative and qualitative research, we should observe strong positive correlations for each pair of methods practices. For example, all qualitative articles should make modest generalizations (practice 12) and choose cases based on theoretical and substantive importance (practice 13), whereas all quantitative articles should generalize broadly and use random samples. Moderate or weak pairwise correlations would undermine the two-cultures hypothesis because it would show that more than a few qualitative articles follow the quantitative culture for a selected practice and that quantitative studies follow the qualitative culture. A test of this implication is useful, regardless of the result of the test of the second observable implication. The two-cultures hypothesis could be incorrect even if we find a single dimension in our data because many different combinations of practices that do not reflect the two cultures can be reduced to a single dimension.[10] In the absence of evidence for a single dimension, a test of the third implication is also interesting because the pairwise correlation might indicate why there is not a single qualitative-quantitative dimension in the data.

Fourth, if the two-cultures hypothesis was true, we should observe that a large proportion of method and research design decisions are in line with the expectation. This general implication can be specified for the incidence of *one practice over multiple quantitative and qualitative articles* and for the prevalence of *all 25 practices within one qualitative or quantitative study*. When we take method practices as the units of analysis, we should, for example, observe that qualitative articles mostly discuss equifinality in their empirical analysis, whereas quantitative articles do not (Goertz and Mahoney 2012, chapter 6). When our unit of analysis is the individual empirical study, we should observe that the 25 method and design decisions largely correspond to the expected culture. For example, we should observe a large share of qualitative articles that explain the outcome of a single case; analyze causal

mechanisms; and so on (Goertz and Mahoney 2012, chapter 19). Unless both observable implications are fully correct and the observed proportion of expected practices is 1 across and within articles, there is added value to testing both implications empirically. Suppose we find that a qualitative practice is implemented in 80% of all qualitative articles under analysis. This would appear to serve as confirming evidence for the two-cultures hypothesis. However, a share of 0.80 for each practice, for example, allows for the possibility that only 60% of the articles display all the qualitative practices.[11] Since we cannot infer the incidence of culture-conforming practices within one article from their prevalence across articles, and vice versa, we formulate implications separately for individual practices and articles.

Both observable implications are important to test regardless of the finding for the third implication. Strong positive correlations between practices are not necessarily evidence of the presence of the expected cultures. A large number of patterns in qualitative and quantitative methods applications can give rise to strong positive correlations and still be in discord with the presence of the two cultures. If we find moderate or weak pairwise correlations, an analysis of the fourth implication would be of value because the proportions of methods practices in qualitative and quantitative work would show why a correlation deviates from the expectations.

Fifth, following the idea of three waves of qualitative methods development (Goertz and Mahoney 2013b), we should expect the qualitative culture to become more pronounced over time (Goertz and Mahoney 2012: 226). Before the publication of *Designing Social Inquiry (DSI)* (1994), qualitative methods primarily centered on "the comparative method" (Collier 1993; Hall 2008: 308). The focus was on what is now called the cross-case level and causation was inferred via symmetric associations between cause and effect through Mill's methods (Lijphart 1971). Until the mid-1990s, notions of "set theory", "process tracing" and "causal mechanism" were largely absent. This implies that qualitative research should not consistently reflect the qualitative culture during the first wave.[12]

The publication of *DSI* inspired a new wave of work on qualitative methods, which pitted a set-theoretic and asymmetric view against a correlational, symmetric conception of causality. It also shifted attention from comparisons on the cross-case level to process tracing and the analysis of causal mechanisms (Beach and Pedersen 2013a, chap. 2; George and Bennett 2005; Mahoney 2010). These lines of thinking were developed in the 2000s and needed time to influence the practice of qualitative research. The implication is that qualitative articles should start reflecting the qualitative culture more in

the early 2000s and most strongly around 2010 (our main period of analysis ends in 2012, when ATTC was published).

In contrast, we expect the quantitative culture to be largely invariant over time. Quantitative methods changed substantially over the past 20 years because of an increased interest in causal identification and Bayesian statistics, and the development of new estimation techniques for different forms of data such as panel or multilevel data (Keele 2015; Lebo and Weber 2015). From the perspective of ATTC, however, these developments all occurred within an already established framework of quantitative research and are not related to the constitutive elements of the quantitative culture. We expect to find evidence for a quantitative culture at any point during our research period and little variation in methods applications over time. Taking our expectations as a whole, it follows that qualitative and quantitative methods practices should become more dissimilar over time because of a reorientation of qualitative methods. Table 1 summarizes the five observable implications that we will test empirically in the remainder of this paper.

These observable implications are not equally relevant to the empirical test of the two-cultures hypothesis. Observable implications 1 (high validity and empirical relevance), 2 (a single dimension of methods practices), and 4 (high shares of predicted practices) are, in our view, implications that directly follow from the two-cultures hypothesis and should be confirmed. We pay particular attention to the empirical results for these three implications in the following. Observable implications 3 (strong pairwise correlations) and 5 (changes over time), in contrast, can be conceived of as indirect implications of the two-cultures argument. They should also find empirical confirmation if the two-cultures hypothesis holds, but we take them to be less relevant in the context of our analysis.

Empirical Strategy

Article Collection[13]

We test the observable implications against a stratified random sample of 180 empirical research articles. In line with the perspective taken in ATTC, we limit the selection of articles to those answering causal research questions and include neither purely descriptive articles, nor those based on a interpretivist framework into the analysis (Goertz and Mahoney 2012, chapter 1). The sample of 180 articles is stratified on three levels. First, we stratify by the method that is employed in the article, selecting 90 articles that apply a qualitative method and 90 that implement a quantitative method. The 180 articles

Table 1. Summary of Observable Implications.

OI	Expectation	Specification of expectation
1	High validity and empirical relevance	All 25 practices used to describe differences between the two cultures are empirically relevant practices in qualitative and quantitative research.
2	Single dimension	Methods practices can be reduced to a single "methods culture" dimension.
3	Strong pairwise correlations	Individual practices should be strongly positively correlated with each other.
4	High shares of predicted practices	The proportion of qualitative and quantitative studies that display the expected methods practice should be high for individual practices. The proportion of expected practices should be high for individual articles.
5	Differences between quantitative and qualitative methods should become more pronounced over time.	Quantitative work consistently reflects the quantitative culture.

were randomly sampled from *journal-periods*, capturing that we additionally stratified articles by journal on a second level and by time period on a third level. We chose Anglo-American and European outlets that cover all the major sub-disciplines of political science:[14] *American Journal of Political Science (AJPS)*, *American Political Science Review (APSR), Comparative Political Studies (CPS), European Journal of Political Research (EJPR), International Organization (IO)* and *World Politics (WP).* For each method and journal, we distinguished three time periods to test the fifth observable implication. We expect a trend towards greater coherence within the two methods cultures over time, and especially in qualitative research. The first period, 1990–1994, captures the time before the publication of *DSI* and the dominance of "the comparative method" in qualitative research. The period 2000–2004 is an interim phase in which we might already see instances of new developments in qualitative methods covering process tracing, set theory, etc. The third period, 2008–2012, was chosen because it is most likely to show evidence for a qualitative culture (see previous section). Altogether, we selected ten articles per journal-period (five qualitative and five quantitative), which yielded 30 articles per journal, 60 articles per period, and 180 articles in total.[15]

We chose these journals for two reasons. First, the snap-shot analysis in the appendix of ATTC draws on APSR, CPS, IO and WP as the empirical basis for measuring the incidence of quantitative and qualitative research and specific methods. This makes these journals a straightforward choice for a test of the two-cultures hypothesis.[16] Second, these journals are considered premier outlets in political science across our period of analysis (Giles and Garand 2007). If two methods cultures exist, we should find evidence for them in journals that are arguably paradigmatic for setting the methods standard in the discipline. Consequently, we find it most likely to identify the expected method patterns in these journals.

Despite these arguments in favor of the chosen journals, there might be the concern that the six journals or a subset of them are least likely to produce confirming evidence for the two-cultures hypothesis. The concern could be rooted in the fact that two of them (AJPS, APSR) mostly publish quantitative work and that they or others might attract work following a "quantitative, correlational template" that is in discord with the qualitative culture. However, we find it conceptually difficult to argue that a qualitative culture exists except for articles published in any set of six journals——and ATTC does not make such an argument. We discuss the potential bias introduced by the selection of these journals as well as the details of the sampling procedure in Appendices B–D.

Coding of Method Practices

The coding of method practices in the 180 articles proceeded as follows. We first developed a coding scheme and instructions based on Chapter 18 of ATTC and the chapters dealing with one or more methods practices (see Appendix A). Guided by the coding scheme, each of the authors independently read and coded all 180 articles on each of the 25 items capturing a specific methods practice.[17] We coded an item "0" when it corresponded with the expectations of the quantitative culture and "1" when it conformed to the qualitative culture.[18]

As an example, consider item 4 ("process tracing"), which asks whether an article includes the empirical analysis of an historical process in at least one case. We coded the item "1" if a process was reconstructed in the empirical analysis that went beyond reporting isolated within-case evidence. Hsueh's (2012) analysis of differences in the reregulation of economic sectors in China and India, which was sampled as one of the qualitative articles published in CPS in the third journal period (2008–2012), includes a detailed historical description of the reregulation schemes in both countries in the textile

and telecommunication sectors, each spanning multiple pages. This was coded "1". However, we also coded less detailed and extensive process narratives as conforming to the qualitative culture. Weyland's (2010) article on the diffusion of anti-regime protests in the 1830–1940 period, drawn from the same journal-period as Hsueh's, only provides a few sentences of qualitative evidence for each of the many cases discussed in the article. Nonetheless, all present a clear, albeit brief, historical narrative on how external influence affected the chances for successful pro-democratic mass mobilization in Europe. Hence, we coded the process tracing item as conforming to the qualitative culture. As an example of a qualitative article that in our view did not include the empirical tracing of processes, consider Hopf (1991). The article tests Waltz's theory on the impact of multi- vs. bipolarity in the international balance of power on the stability of the international order. While the article provides a vast amount of historical data on the power distribution in the 16th century, it does not include systematic process narratives as to how these measures of state power affected stability or instability in individual cases. Hence, we coded the process tracing item 4 as conforming to the quantitative culture, i.e., "0".

We assigned a code of "99" (interpreted as "missing") when an empirical article followed a practice not predicted by ATTC, which is possible for items for which the qualitative and quantitative practices outlined in ATTC are not jointly exhaustive. For example, the indicator capturing the rules of case selection (item 13) was coded "99" when an empirical study did not explain why the cases were chosen for analysis, as this reflected neither the quantitative nor the qualitative culture. Another use of the "99" code occurred when a method practice was simply absent in an empirical article. For example, item 21 asks for *how* a typology is constructed, which requires *that* an article features a typology in the first place (most articles do not; see Section "Observable Implication 1: Relevance of Method Practices"). The possibility that the meaning of the "99" code differs across items and articles is not problematic for our analysis.[19] We are specifically interested in the incidence of methods practices conforming to the quantitative culture, the qualitative culture or neither, and not in a general mapping of design and method decisions in political science.

The intercoder reliability for all articles and items that we coded has a Krippendorff's alpha of 0.77, which is slightly below the conventional threshold of acceptability of 0.80.[20] After each coder had analyzed all articles, we discussed the items with different codes to decide on what code best aligned with the item. For many items, agreeing on a final code was not difficult because the deliberation and re-reading of the article convinced one coder

that the other's coding decision was more plausible. When coder agreement was difficult to achieve, which usually occurred when the empirical study was ambiguous in regard to a methods practice, we coded conservatively in favor of the culture. In that event, we assigned a "1" to the item in question for a qualitative article and a "0" when it was a quantitative article.[21] As an example, consider the divergent coding of Traxler's (1992) qualitative analysis of the determinants of state policy in Austria regarding the question of whether the article follows the qualitative or quantitative culture in the ontology of concepts (item 18). One of us read the conceptual discussion as including a discussion of the concept's defining characteristics and their relations and, in consequence, as following the qualitative culture (Goertz and Mahoney 2012: 128). The other saw the conceptual discussion as closer to what ATTC considers a "focus on issues of data and measurement, and less on semantics and meaning" (ibid.) and, thus, corresponding to the quantitative culture. After discussing our positions, we agreed to disagree and coded the item "1", i.e., in favor of the culture.

The spreadsheet containing all sampled articles, coder-specific and agreed upon codes is available in the online repository.[22]

Observable Implication 1: Relevance of Method Practices

The starting point of the tests of observable implication 1 is the screening of the definition of all qualitative and quantitative practices; this allows us to separate the items that are necessarily relevant from those that might not be important in practice. A method practice is relevant when the qualitative and quantitative practices are jointly exhaustive and an empirical study has to follow one or the other. We identify six items as potentially invalid indicators of method cultures because the corresponding practices might not be observed in practice: 5 (counterfactuals), 9 (counterfactuals), 16 (triangular data), 20 (data transformation), 21 (typologies) and 24 (counterfactuals). All other items (except item 13, case selection, see below), cannot be coded as "missing".

We assess the conceptual validity and empirical relevance of each of the six methods practices by its incidence in empirical research. For the 90 quantitative and qualitative articles, we calculate the share of studies that do *not* display a methods practice and are coded "99". We split the sample into the quantitative and qualitative part to detect possible differences in relevance between methods. In the absence of a convention to draw on, we find it

reasonable to argue that, to be designated as relevant, a methods practice should at least be present in more than 50% of all empirical studies. Figure 1 presents the shares of missings for the six items that can be coded "99".

All six practices for which a "missing" code of "99" was possible have missings in more than half of all qualitative and quantitative articles. We observe variation in the extent of missingness, ranging from complete irrelevance in qualitative articles (understanding of triangular data, item 16, and data transformation, item 20) to the absence of practices in about 70% of all articles (counterfactuals in quantitative articles; items 5, 9, 24). Using our criterion of empirical importance, we observe that the six possibly irrelevant practices are not valid characteristics of the two cultures because they are rarely observed in practice. If there are qualitative or quantitative methods cultures, they are constituted by the subset of practices that are relevant to empirical work. For this reason, we limit the following analysis to the empirically important items unless we explicitly state otherwise. In Figure 1, we left aside item 13 (case selection) because the "99" code denotes the non-discussion of how cases were chosen *or* that a case selection strategy other

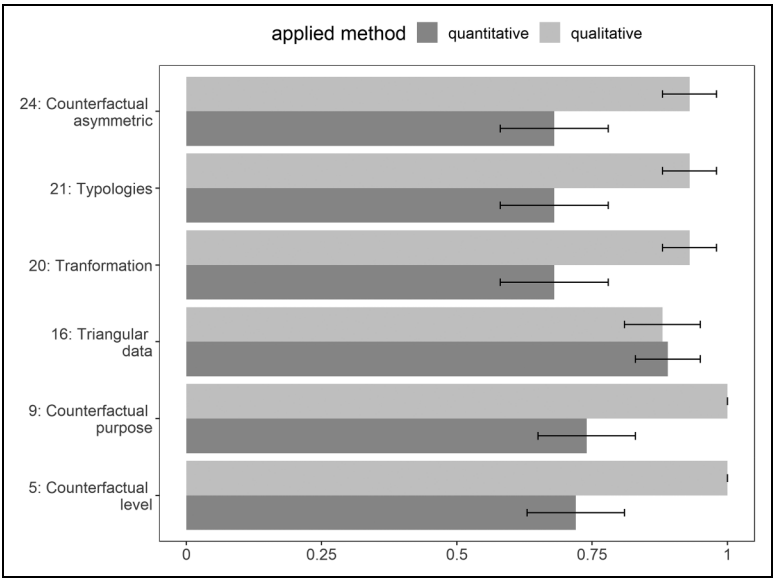

Figure 1. Proportion of missings and 95% confidence interval.

than those discussed in ATTC was followed. A separate analysis of case selection strategies shows that qualitative articles follow the qualitative culture less often than quantitative articles and that both show a sizeable share of "99" codes. 34% of all quantitative articles do not sample cases randomly [95% confidence interval: 0.24–0.34] and 66% of all qualitative articles do not display the expected practice [0.56–0.76].

Observable Implication 2: A Single Dimension of Method Application

We run a confirmatory factor analysis (CFA) to test the second observable implication that the realization of qualitative and quantitative methods represents a single underlying dimension. The data for the 18 relevant items that we include in this analysis are categorical and do not follow a multivariate normal distribution, which is an assumption in CFA.[23] We take this into account by using a robust weighted least squares estimator for testing the null hypothesis that a single dimension exists. Table 2 summarizes the results.

The Pearson χ^2-test for model misspecification is significant, allowing us to reject the null hypothesis that the data represent a single methods dimension. A second test is the root mean square of approximation (RMSEA) test that evaluates the degree of deviation from perfect model fit for a single dimension. According to the conventional interpretation, values below 0.1 are taken as acceptable and values below 0.05 are good (Hu and Bentler 1995). The RMSEA test for the existence of a single dimension yields 0.128, with a 90% confidence interval from .113 to .144. It fails to meet the standard benchmark for acceptable model fit. These findings are confirmed by the Lewis-Tucker index, which is an alternative test to the RMSEA. The estimate is 0.90, which fails to meet the conventional threshold of 0.95 that signals good model fit. In sum, the three measures are not far from the conventional benchmarks indicating good model fit, but they fail to meet any of those benchmarks, thus indicating that there is not a single methods dimension.

Table 2. Fit Measures for Confirmatory Factor Analysis of 18 Items.

measure of fit (robust estimation)	Value
$\chi^2_{(df = 135)}$	534 ($p < .001$)
RMSEA	0.128 (90% CI: 0.113–0.144)
Lewis-Tucker index	0.90

Observable Implication 3: Strong Pairwise Correlations of Items

The third implication predicts that we should observe strong correlations between all 18 empirically relevant items because we would consistently code quantitative practices as "0" and qualitative articles as "1".[24] We calculate Cramér's V and the 95% confidence interval for all pairwise correlations to test this implication (Figure 2).[25] We apply a permissive benchmark and designate a correlation as high if the correlation coefficient is larger than 0.5.

Figure 2 shows that 10 of the 151 pairwise correlations are at 0.75 or higher and that 14 are clearly above 0.5. The largest share of correlations is in the medium and lower ranges, with 91 correlations being below 0.25, which is also reflected in a median correlation of less than 0.25.[26] The pairs of items that correlate with each other at a level higher than 0.5 mainly come from two sets of items. First, items 1, 2, 3 and 4, which capture different method and design decisions about how individual cases are treated, are strongly correlated with each other. The other set of items comes from different dimensions and can be summarized as those related to asymmetric relations in the data, with ten items in total (items 6, 8, 11, 14, 17, 18, 19, 22, 23, 25). Some of these correlations are to be expected because items 22, 23, and 25 (symmetry/asymmetry of concepts and causal inferences) must all be coded either "0" or "1". Altogether, we conclude that the third observable implication receives little empirical support.

The findings for observable implication 2 and 3 suggest that the two-cultures hypothesis does not hold up against empirical evidence. Based on these insights, we find it necessary to take a closer look at the data. From this point on, we disaggregate the pooled data into the quantitative and qualitative subsample and test the implications separately against the two subsamples. This allows us to determine whether the reasons for the differences between expectations and findings lie with the quantitative or qualitative articles or both.

In Figure 3, we present the pairwise correlations separately for the quantitative and the qualitative articles.[27] We observe a difference between the two groups of articles because the correlations are, at least to some degree, in the upper range for quantitative work and mainly in the lower range for qualitative articles. The median correlation is about 0.7 for quantitative articles and 0.1 for the qualitative articles. Among the pairwise correlations for the quantitative articles, items 7 (interactions), 10 (equifinality), 12 (scope of inference) and 18 (ontology of concepts) stand out because they are only weakly correlated to all other items. For the qualitative articles, in

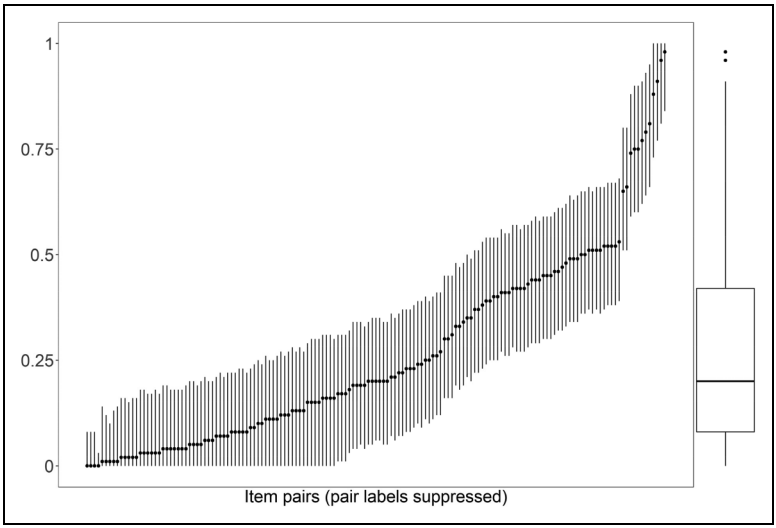

Figure 2. Pairwise correlations of individual items (Cramér's V) with 95% confidence intervals.

contrast, only a handful of items achieve a level of correlation that is higher than the median for quantitative work. The analysis of the third observable implication clearly indicates that quantitative and qualitative research do not fully follow the culture and that the difference between expectation and evidence is much larger for qualitative than for quantitative articles.

Observable Implication 4: High Shares of Culture-Conforming Applications

We first assess the implication for the share of culture-conforming practices by taking *items as the unit of analysis*. For each item, we calculate the share of how often it is realized in accord with the corresponding culture.[28] We present the data separately for qualitative and quantitative articles to detect possible differences between them.[29] We use a share of 0.75 as the threshold for evaluating the incidence of method practices. We infer that a method practice confirms the observable implication if the share is larger than 0.75 and as not confirming it otherwise.

Figure 4 shows that, for quantitative articles, 15 practices out of 18 are statistically indistinguishable from 1 and support the expectation of a quantitative

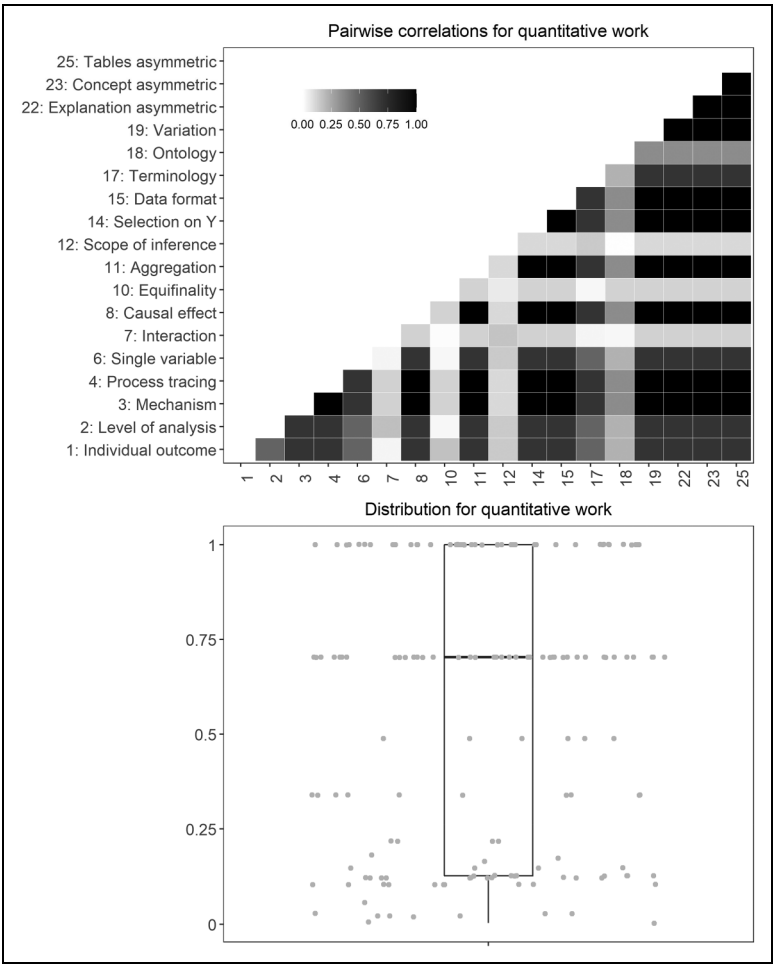

Figure 3. Pairwise correlations by type of applied method. (continued)

methods culture. The three practices that stand apart relate to items 7 (interaction effects), 10 (equifinality) and 12 (scope of inference) that are statistically indistinguishable from the share of qualitative practices. The results show that, for item 7, about 40% of the quantitative articles estimate an interaction effect; for item 10, about 45% of the quantitative articles discuss equifinality; and, for item 12, only 50% of quantitative articles make a broad generalization claim.

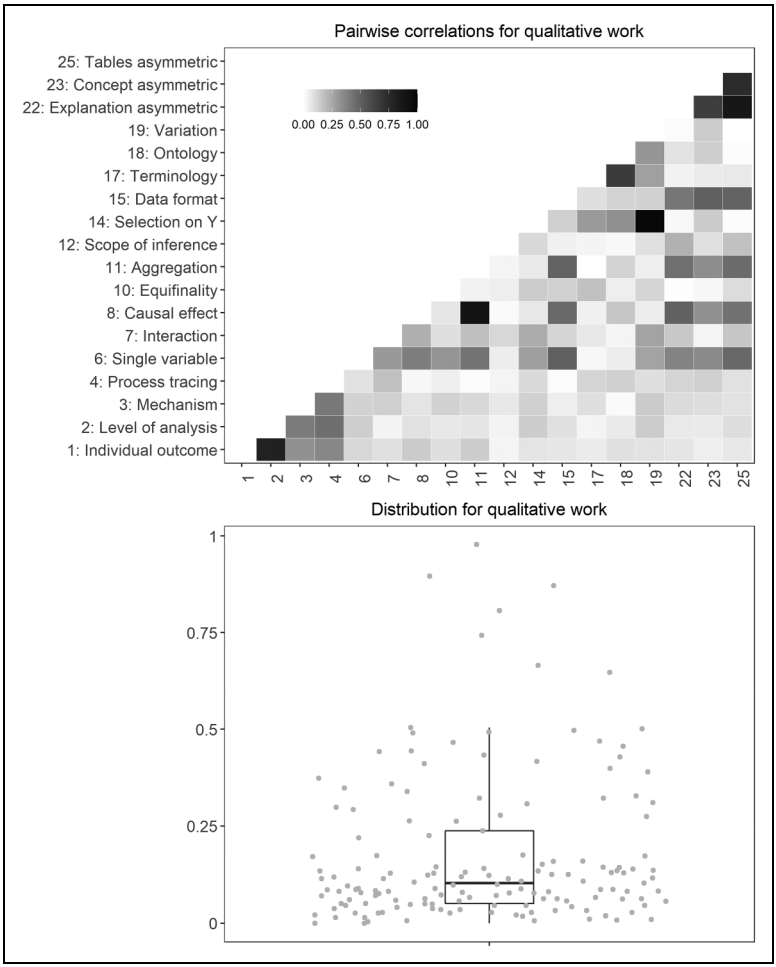

Figure 3. Continued.

The findings differ for qualitative articles wherein large proportions of culture-conforming codes can only be identified for the first dimension (items 1–5), which relates to the way in which individual cases are handled. 14 practices out of 18 are either as much quantitative as they are qualitative (confidence interval includes 0.5) or are closer to the quantitative culture. This includes items 22, 23 and 25 that capture whether the analysis is symmetric or asymmetric and indicate that qualitative articles usually take a

symmetric view. In sum, the findings for this implication support the divergent findings for the third implication. The moderate to weak correlations of items for qualitative articles have their source in the heterogeneous implementation of methods practices in qualitative research.

The second test of the fourth implication takes *articles as the unit of analysis* and calculates the share of the 18 practices that an article comprises in accord with the corresponding culture. In Figure 5, the left distribution shows that quantitative method practices are mostly in line with the corresponding culture for all 90 articles. Whereas this is less surprising after having seen Figure 4, Figure 5 shows that qualitative articles have a higher degree of heterogeneity.

For the qualitative culture, we observe a much wider dispersion of practices, with the third quartile being at a share of culture-conforming practices of 50%. Only a handful of articles reach a share of culture-consistent method

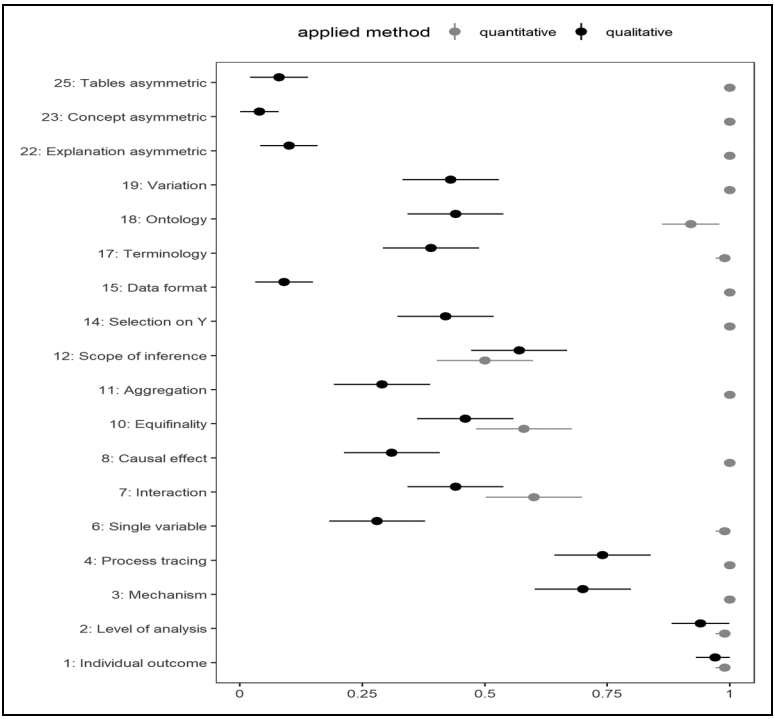

Figure 4. Proportions of culture-conforming codes and 95% confidence intervals.

practices of about 80%, which is the median for quantitative work. In comparison, there are more articles that are more quantitative in nature, with a share of qualitative practices of less than 0.25, which means more than 75% do follow the unexpected, quantitative culture. This confirms the mixed conclusions we made before and indicates that the two-cultures hypothesis is principally valid for quantitative articles but has little validity for qualitative articles.

Observable Implication 5: Method Applications Over Time

The fifth implication is that the differences between quantitative and qualitative approaches should become more pronounced over time as qualitative methods become more developed and formalized. Based on the results in the previous section, the most direct and, in our view, meaningful evaluation of this implication is to determine whether the proportion of articles corresponding with the two-cultures hypothesis increases over time (Figure 6).

For quantitative articles, we see a very stable pattern across the three periods. The only change that stands out is an unexpected increase in the

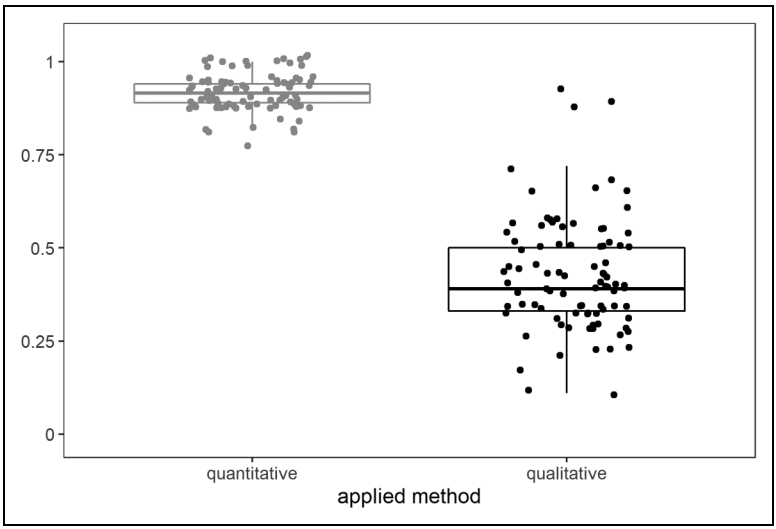

Figure 5. Proportion of culture-conforming practices per article by applied method.

share of articles that estimate an interaction effect from 2000–2004 to 2008–2012 (item 7). Overall, Figure 6 confirms our expectation that there should be little temporal variance for quantitative methods practices.

The development of qualitative methods practices over the three periods does not confirm our expectation of an increasingly pronounced qualitative culture. This implies that neither method became *more* distinct in comparison. The three panels indicate that the variance across qualitative practices decreases over time.[30] Method practices for 2008–2012 show that there are three clusters in the most recent period (item 15, capturing the format in which data are presented, is not part of a cluster of practices): first, one cluster with items from dimension 5 (items 22, 23, 25 which capture symmetry/asymmetry of concepts and causal arguments) that follow the quantitative culture; second, one cluster representing dimension 1 (items 1 to 4, which indicate how individual cases are treated), with items 3 (empirical analysis of causal mechanisms) and 4 (tracing of empirical processes) showing a trend toward becoming more qualitative over time and largely following the qualitative culture in terms of levels; and third, one "neither-nor" cluster that includes ten items from different dimensions that converge in the range of about 0.25 to 0.5 over time. Overall, there is no evidence for a strengthening qualitative culture over time, including the most-likely period 2008–2012.

Summary and Alternative Explanations

The empirical evaluation of the five observable implications provides little support for the two-cultures hypothesis (Table 3). Some practices in quantitative research are as qualitative as they are quantitative, but the overall level to which quantitative articles conform to the quantitative culture is high. This is very different for qualitative research wherein articles show a much greater degree of diversity in how design and method decisions are made and only a handful of practices meet the expected ideal of the qualitative culture. Taken together, the main reason that the two-cultures hypothesis is not confirmed rests in the misrepresentation of qualitative empirical research.[31]

A look at our findings suggests that a central reason for qualitative articles differing from the expected culture is the underlying assumption inherent in ATTC that qualitative research is based on the foundation of set theory. Our analysis shows that are qualitative research is based on neither a set-theoretic modeling and handling of data (captured by items 15, 16, 23, 25), nor on set-relational and asymmetric modeling of causality (items 6, 8, 11, 22).

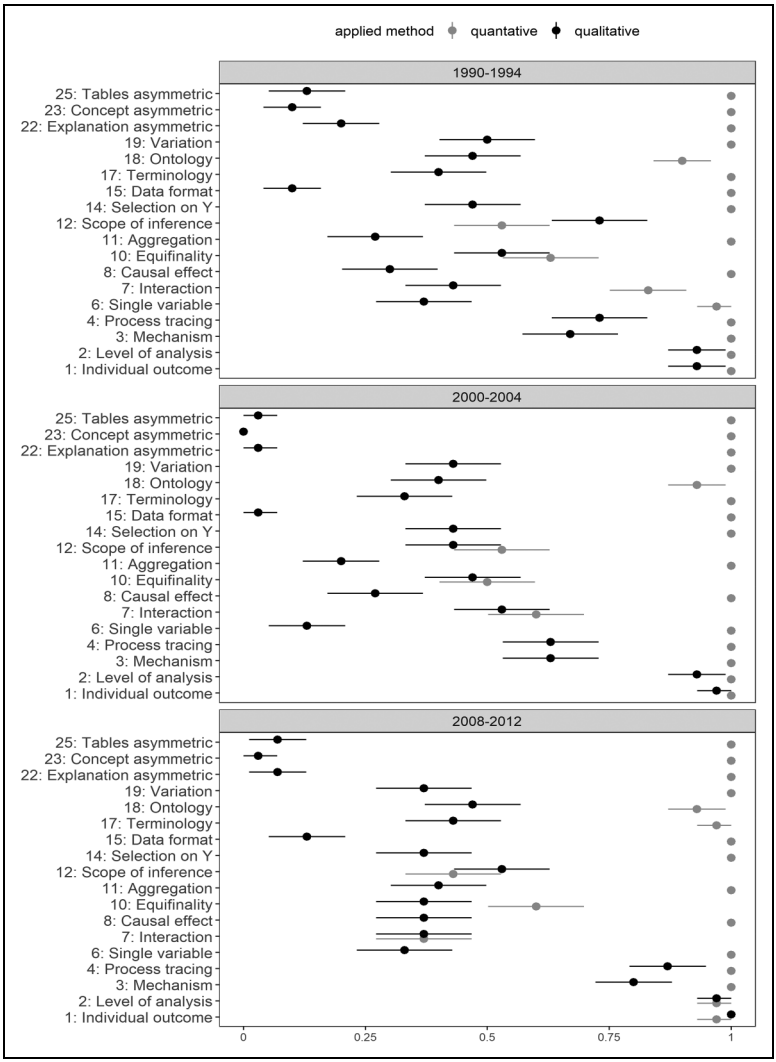

Figure 6. Proportions of culture-conforming codes and 95% confidence intervals for subperiods.

Some of these items work well as valid descriptors of empirical research using Qualitative Comparative Analysis (QCA, Marx et al. 2014), but it seems that they cannot be successfully exported to qualitative empirical

research per se (see also the results of a cluster analysis of qualitative practices in Appendix E.3.2).

There are be two alternative explanations for our findings other than that many qualitative articles do not follow the qualitative culture. First, qualitative research could be characterized by "implicit but quite common practices" (Goertz and Mahoney 2012: 9, 11) that are not amenable for identification through a standardized coding and evaluation process as applied in this article. If this was true, our disconfirming findings for qualitative research could be explained by having misread and miscoded qualitative practices. There are two reasons why we are not convinced by this alternative. The first is that, according to ATTC, despite being implicit, qualitative methods characteristics should be "readily identifiable" and can be "reconstructed" by "a broad reading of qualitative studies, including an effort at systematically coding qualitative research articles" (Goertz and Mahoney 2012: 2, 7). Moreover, whenever we were unsure about the coding, we coded conservatively in favor of the culture. This should offset at least to some degree the potential impact of ambiguousness due to implicit use of certain methods practices on our results. The second reason for our skepticism is that only a small subset of the 25 practices that we analyzed can be realized without making them explicit. In our reading, only six items could *potentially* capture implicit practices and lead to ambiguous codings.[32] If the implicit-practices counterargument was true and if we have miscoded method and design decisions[33], this might change some of the findings we have presented, but would not overturn the overall conclusion of heterogeneous qualitative research practices.

The second alternative explanation is that the two-cultures hypothesis is empirically true for methods practices in political analysis, but that our data are biased against uncovering them.[34] Our data collection could be biased for multiple reasons. One source of bias could be that we evaluate journal articles and not books. [35] It is easier to present the *evidence and results* of a qualitative study in a book. This advantage of books only concerns items 3 and 4 about mechanisms and processes in the empirical analysis. Our evidence showed that these practices have been largely in accord with the qualitative culture, so the focus on articles does not seem to be a problem. For all other practices, the type of publication should be unrelated to the *implementation* and of qualitative methods because this does not require extensive discussion. The justification of design and method decisions should require about the same space in a qualitative study as in a quantitative study. Those articles in our data that do largely correspond to the qualitative culture (Figure 5) demonstrate that one can follow the qualitative culture in journal articles.

Table 3. Results of the Empirical Tests of Five Observable Implications.

OI	Expectation	Results	OI corroborated
1	High validity and empirical relevance	Six out of the 25 practices that could not be relevant (no exhaustiveness of expected quantitative and qualitative practice) are not relevant in empirical research.	No
2	Single dimension	Method practices do not reflect a single dimension of method practices.	No
3	Strong pairwise correlations	Most method practices do not correlate strongly with each other. Some of those that do, do so by the way they are defined in ATTC.	No
4.	High shares of predicted practices	The proportion of items in line with the used method is high for quantitative articles and low for qualitative articles.	Yes (quant) and no (qual)
5.	Strengthening qualitative culture over time, constantly strong quantitative culture	Strong conformity with quantitative culture over time (with few exceptions). Most qualitative practices and articles are not following the qualitative culture. No discernible trend towards qualitative culture over time.	Yes (quant) and no (qual)

A second potential source of bias could be the journals that we have chosen. The editorial policy of journals might be predisposed in favor of quantitative work or a certain type of research that is inspired by the quantitative template more generally (such as in King et al. 1994). There might also be regional biases, as five out of the six journals might follow an Anglo-American tradition that has traditionally been dominated by quantitative research and "symmetric" understandings of causality. In comparison, European political science has been described as being more open to qualitative research and pluralism (Moses et al. 2005). If this alternative explanation

was true, it would imply that articles following the qualitative culture have a low probability of getting published and were published in journals we did not select; or that they are revised throughout the editorial process to better conform to the quantitative culture; or that authors expect such preferences for the quantitative culture and refrain from submitting to these journals or follow the quantitative culture preemptively.

We try to assess journal-induced bias, broadly speaking, by comparing methods practices across the six journals. Figure 7 presents the proportions of practices that conform to the corresponding culture separately for quantitative and qualitative articles and each journal. We observe that the pattern found for the pooled data (Figure 5) holds in very similar ways across all journals. IO and WP have a somewhat larger dispersion of qualitative articles, suggesting that there is no regional bias because the articles in these two journals are "more qualitative" than those sampled from EJPR. We also think that the panels in Figure 7 do not support the conclusion that editorial policies account for a suppression of the qualitative culture. These journals are known for having a different orientation—AJPS and APSR being more quantitatively oriented than the others—and the data for each journal pool observations from different editorships who are likely to follow different editorial policies.[36]

Figure 7 does not conclusively refute the possibility that the choice of journals might be biased against the qualitative culture. However, we believe that the presented evidence indicates that journal bias is not driving the empirical results. This is confirmed by the results of 30 additional qualitative articles published in 2018 in a large number of different journals other than the six that underlie the main analysis (see Appendix E.3).

Finally, our analysis could be problematic because it might end too early. This would be a valid claim if empirical articles only picked up on recent developments in qualitative methods after 2012. We address this possibility by coding 30 additional qualitative journal articles published in 2018 and in 28 journals other than the six main journals. We present the sampling procedure and detailed results in Appendix E.3. We test observable implication 4 (share of culture-conforming method practices) against the 30 articles because the disaggregated perspective makes it easiest to see where the 30 additional articles differ from the 90 articles in the main analysis. We fail to find meaningful differences between the two samples when taking items and articles as the unit of analysis. These findings strengthen our main results and indicate that they are neither conditional on the chosen journals nor on the three periods of analysis and that there is no qualitative culture that exists exclusively outside of the six premier journals included in the main dataset.

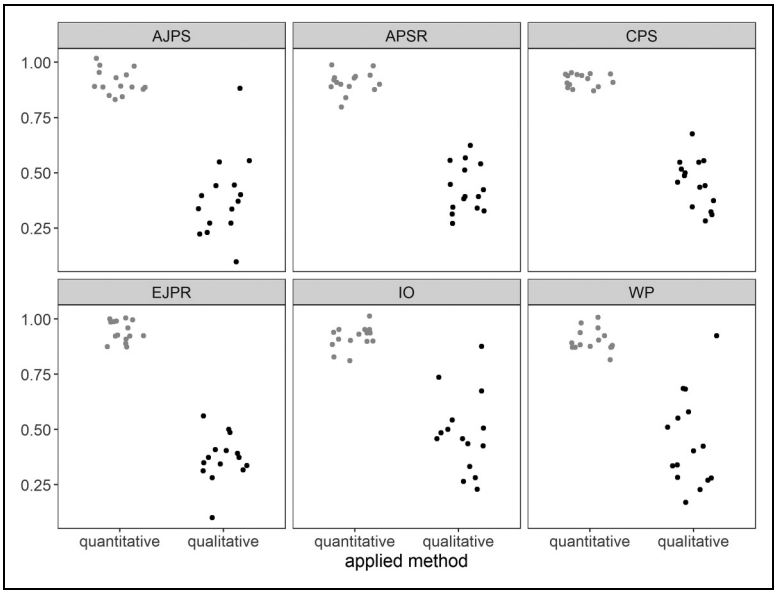

Figure 7. Proportion of culture-conforming practices per article per journal.

Discussion

In this paper, we take the seminal hypothesis of "A Tale of Two Cultures" about the presence of two coherent and distinct method cultures and derive five observable implications from it. The implications allow us to perform a comprehensive empirical test indicating that the two-cultures hypothesis is not an empirically valid description of common methods practices in political science. The nuanced picture that we obtained from a stratified random sample of 180 empirical articles is that, with a small number of significant exceptions such as the analysis of interaction effects, the implementation of quantitative methods follows the expected culture. Qualitative methods, in contrast, neither conform to the qualitative culture nor do they show a clear trend to become "more qualitative" (as per the two-cultures hypothesis) over time, including the period 2008–2012, which is identified as a most-likely time frame. Since qualitative methods continue to be discussed along the dimensions of asymmetry, mechanisms, process tracing and set theory, it is possible that empirical research might reflect these characteristics more clearly in the future and that the height of the "qualitative culture" is yet

to come. While we cannot forecast future method developments, our analysis of 30 randomly selected qualitative articles published in 2018 casts doubt on this argument (see Appendix E.3). Overall, our analysis shows that qualitative research is not mainly based on the set-theoretic underpinnings stipulated by ATTC. A set-theoretic framework for analysis with a distinct mode of causal inference is valuable and can be used in qualitative research. However, our results suggest that a set-theoretic perspective is not all there is to qualitative research and that a certain share of articles adopts a "symmetric", difference-making perspective on causal inference (broadly understood) as it underlies quantitative research.

If qualitative research does not conform to the qualitative culture as defined in ATTC, the follow-up question is whether our data suggest the existence of other patterns in qualitative practices. We address the question by running an exploratory cluster analysis on all 90 qualitative articles in our sample (see Appendix E.4 for the full procedure and results). In this, we draw on Koivu and Damman's (2015) conceptual distinction between three approaches within qualitative research to specify three clusters *ex ante*: "quantitative emulation (QE)", "eclectic pragmatism (EP)", and "set-theoretic approaches (ST)".[37] According to Koivu and Kimball Damman, these three approaches are based on different foundations that lead to different method and design decisions in causally oriented qualitative research. The analysis shows that the vast majority of qualitative articles in our sample are grouped into two clusters: one containing 48 articles (53.3% of all qualitative articles) and one containing 35 articles (38.9%). A third cluster is small and contains only seven articles (7.8%).

To determine whether the three clusters correspond to the three conceptual groups suggested by Koivu and Damman, we calculate the per-item means within each cluster and compare them across the three clusters: the closer the value is to 0, the closer the cluster comes to the quantitative ideal; the closer the value is to 1, the closer the cluster is to the set-theoretic qualitative ideal described in ATTC. The first cluster has a median value of 0.25; it thus corresponds well to Koivu and Kimball Damman's QE label—qualitative research that broadly emulates the quantitative template. This includes an interest in the individual effect of individual variables and treating causal relationships as symmetric. The median of the second largest cluster is close to 0.5. This cluster thus corresponds to the EP category as articles in this cluster on average show an eclectic mixing of quantitative and qualitative practices. The smallest cluster, which includes only seven individual articles, has a median share of culture-conforming practices of 0.79. The seven articles in the third cluster are well-described by Koivu and Kimball Damman's ST

label——and the set-theoretically defined qualitative culture according to ATTC. However, and echoing Koivu and Kimball Damman's insight, our empirical analysis shows that the ST approach is but one of numerous "varieties" of qualitative research——and is not what the majority of qualitative researchers does.

Future research on the practice of empirical political research could proceed in four directions. First, there is room for conceptually defining more valid and empirically relevant indicators for mapping method practices in political science. Second, for the reason previously discussed, it would be useful to run a follow-up study to test for a stronger presence of the qualitative culture and to measure a potential prescriptive impact of ATTC on qualitative methods practices since 2012. Third, although we don't believe that the selection of the six journals biases the results against the two-cultures hypothesis, it would be empirically insightful to analyze articles and possibly books covering a variety of substantive and empirical foci. This would broaden the empirical basis and could serve as a replication of the present analysis to deepen the insights and method practices in political science. Fourth, future research should address and further investigate the existence and causes of systematic differences in the application of methods, especially within the broad category of "qualitative research".

Acknowledgments

For excellent research assistance, we are grateful to Kevin Benger, Charlotte Brommer-Wierig, Nancy Deyo, Nils Holzbecher and Nils Hungerland. Earlier versions of this paper have been presented at the APSA Annual Meeting 2017, at the Research Seminar of the Cologne Center for Comparative Politics, and at the German Institute for Global and Area Studies. We appreciate the constructive feedback from the reviewers and Felix Elwert.

Authors' Note

Authors in alphabetical order, both authors contributed equally. All supplementary material can be accessed from the corresponding Open Science Framework repository: https://doi.org/10.17605/OSF.IO/TNKWZ

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was funded by the Fritz-von-Thyssen Foundation (grant number 20.14.0.058). We are grateful for the support of the Fritz-von-Thyssen Foundation.

ORCID iDs

David Kuehn https://orcid.org/0000-0003-2593-0194
Ingo Rohlfing https://orcid.org/0000-0001-8715-4771

Supplemental Material

Supplemental material for this article is available online (see https://doi.org/10.17605/OSF.IO/TNKWZ).

Notes

1. One could question whether 'common practices' qualify as a methods culture (Brady 2013: 253; Goertz and Mahoney 2013a: 282). For the sake of clarity, we follow ATTC's terminology.
2. According to Google Scholar, the foundational 2006 article that presented the two-cultures argument for the first time (Mahoney and Goertz 2006) has been cited 1329 times. The ATTC monograph has been cited 1194 times (as of 22 October 2021).
3. We searched Google Scholar for article citations of ATTC in 2016 (search date: 2017-09-12). 2016 is a reasonable choice of year because ATTC was published in 2012. Taking into account review and journal production times, authors of articles published in 2016 should have had the opportunity to read ATTC and know the content. We collected 54 articles (we excluded other formats) and coded them according to whether they reference ATTC with regard to best methods practices, common practice or in a different context (ambiguous meaning of citation or another issue). 11 citations of ATTC are about common practices.
4. Schwemmer and Wieczorek (2020) conducted an automated text analysis of the abstracts of over 8,000 sociology articles. They found that qualitative and quantitative research differ systematically in regard to their preference for scientific practices; research questions; and the use of certain paradigms. These differences increase over time. This analysis hints at differences between qualitative and quantitative research, but does not provide evidence for the two-cultures hypothesis.

5. On a more general level, we take stock of the application of methods and present data on what researchers do when answering causal research questions. The analysis contributes to what is, to studies that have systematically evaluated method practices in the social sciences and its subfields (Bollen et al. 1993; Bunea and Baumgartner 2014; Coppedge and Kuehn 2019; Groeneveld et al. 2014; Munck and Snyder 2007; Pepinsky 2019). An evaluation of method practices in ethnographic research is presented by Abend, Petra and Sauder (2013).
6. 16 citations of ATTC are about best practices (see footnote 4).
7. Readers who are interested in prescription can use the results to compare their individually defined terms of best practice with current practices, as reflected in our data.
8. The journals are the *American Journal of Political Science*, *American Political Science Review*, *Comparative Political Studies*, *European Journal of Political Research*, *International Organization* and *World Politics*. The periods are 1990–1994, 2000–2004 and 2008–2012.
9. This is despite the fact that QCA per se does not play a prominent role in ATTC. We count only ten explicit references to QCA in the monograph.
10. For example, in the extreme case, all qualitative articles follow the quantitative culture and all quantitative articles the qualitative culture. There would be a single dimension in the data, but method and design decisions would be exactly the opposite of what we expected.
11. To illustrate: Suppose we compare two articles across ten practices A, B, C, etc. to J. The first article might display practices A to H and the second article, C to J. They would only share six practices, that is, C to H.
12. The first publication on process tracing that is usually cited is from George and McKeown (1985). However, it took until the mid-2000s for process tracing to become central to qualitative work (Mahoney 2010).
13. This section presents a shortened summary of the sampling procedure. We provide a more in-depth discussion in Appendix C.
14. We broadly distinguish between American politics, comparative politics and international relations as subdisciplines within political science (see Appendix A). The subfields have no substantive relevance for our analysis except our intention to ensure some variation across the subfields. In total, our sample includes 24 articles on American politics (13.33%), 112 on comparative politics (62.22%), and 44 on international relations (24.45%).
15. We are not interested in the journals per se, which would require a larger sample to reduce the uncertainty of the estimates. Our goal is the collection and analysis of a random sample of articles belonging to the population of quality quantitative and qualitative publications.
16. In addition to the four political science journals, the brief analysis in ATTC's appendix includes two sociological journals (*American Journal of Sociology, American Sociological Review*). We replaced those with AJPS and EJPR

because we want to focus on political science and cover more journals from that discipline.

17. From here on, we use "method practice", "design and method decision" and "item" interchangeably.
18. For item 2 (cross-case vs within-case level of analysis), we introduced a code "3" for qualitative research. We expected, correctly, that many qualitative articles combine an integrated cross-case and within-case analysis that we would have to code as "99" if we followed the original scheme as defined in ATTC. We wanted to avoid the loss of too many qualitative articles for item 2 and opted for adding to a third code "3" to see how often this is done in practice. In our empirical analysis, codes "1" and "3" correspond to the qualitative culture. Item 18 (measurement) suffers from a similar problem because an article can follow both quantitative and qualitative practices. When this occurred, we coded the article in favor of the culture (quantitative articles = 0, qualitative articles = 1).
19. For items 5, 9, 16, 20, 21 and 24, a "99" signifies that the issue is not practiced because no counterfactuals are performed (items 5, 9, 24); triangular data are not discussed (16); data are not transformed (20); or the article does not include a typology (21). For the choice of cases (item 13), "99" can have several meanings, e.g., a convenience sample or the absence of any discussion of how cases were chosen.
20. A breakdown of intercoder reliability by the type of method produces an alpha of 0.76 for qualitative articles and of 0.73 for quantitative studies. See appendix G for a more detailed analysis of individual items and articles.
21. Or a "3" for item 2. In the analysis of our data, we take "1" and "3" as evidence for a qualitative practice and collapse the codes.
22. Please see authors' note at the foot of this article for the address of the online repository.
23. In addition to the six items with a large number of missings, we exclude item 13 (case selection) because it has a sizeable share of "99" codes that are not relevant when testing for a single dimension. The Shapiro-Wilk test for multivariate normality rejects the null of normality at .05.
24. We exclude the six empirically irrelevant items discussed in Section "Observable Implication 1: Relevance of Method Practices" and again exclude item 13 because of the high number of qualitative articles coded "99" on this item. We are interested in correlations between substantively interesting codes, "0" and "1", and not in high correlations resulting from "99" codes (such as for item 16 and 20). We exclude correlations of pairs of identical items because it is necessarily 1.
25. Appendix E.1 includes a heat map for the correlation coefficients that links the degree of correlation to a specific pair of practices for the pooled data.
26. We do not report statistical significance because one easily sees statistical significance for pairs that are weakly correlated.

27. We do not present uncertainty estimates here because we want to include information about the pairs of items and because they would grow large and less meaningful with a reduced sample size of 90.
28. Again, we exclude the items with "99" codes because we can either not calculate a proportion because of too many "99" codes or the confidence intervals would be too large and pointless. We present simple cross-tabulations in Appendix E.2.
29. We cut the bounds of the confidence intervals if the upper bound of the confidence is larger than 1 or the lower bound smaller than 0.
30. See Appendix E.2 for slope plots more clearly tracing the development of items over the three periods.
31. The interpretation of our findings does not change if we take into account the difference between first and second order implications that we introduced in Section "Five Observable Implications of the two-Cultures Hypothesis". Of the direct implications of the 'two cultures' hypothesis (OIs 1, 2, 4), two are refuted while OI4 is only corroborated for the subset of quantitative articles. The same is true for the second order expectations: OI2 is refuted while OI5 is only confirmed by quantitative research.
32. Item 8 (correlation vs set theory), item 11 (linear-additive vs set-relational aggregation), item 15 (dataset vs truth table) and items 22, 23 and 25 on symmetry of concepts and causal reasoning.
33. Our personal coding experience is not that practices were implicit, but were most often made explicit.
34. We chose political science journals in a similar manner as in the appendix of ATTC. This does not render our choice correct, but this alternative explanation extends to the analysis of journals in ATTC.
35. This possibility does not seem to be shared by ATTC that is agnostic about the medium of presentation.
36. We could not sample a sufficient number of qualitative articles for selected journal periods (see above and Appendices B-D). The clustering of the data for AJPS and APSR in figure 7 suggests that the sampling from other journals does not create problems.
37. Koivu and Damman (2015) identify an "empirical interpretivist" practice as a fourth variant of qualitative work that is not relevant here because of ATTC's and, thus, our focus on research aimed at identifying causal relations.

References

Abend, G., C. Petre, and M. Sauder. 2013. "Styles of Causal Thought: An Empirical Investigation." *American Journal of Sociology* 119(3):602-54.

Beach, D. and R. B. Pedersen. 2013a. *Process-Tracing Methods*. Ann Arbor: University of Michigan Press.

Beach, D. and R. B. Pedersen. 2013b. "Review of a Tale of Two Cultures." *Newsletter of the APSA Section on Qualitative and Multi-Method Research* 11(1):2-5.

Blatter, J. and M. Haverland. 2013. "Two Cultures and Beyond: A Plea for Three Approaches." *Newsletter of the APSA Section on Qualitative and Multi-Method Research* 11(1):8-11.

Bollen, K. A., B. Entwisle, and A. S. Alderson. 1993. "Macrocomparative Research Methods." *Annual Review of Sociology* 19: 321-5.

Brady, H. E. 2013. "Do Two Research Cultures Imply Two Scientific Paradigms?" *Comparative Political Studies* 46(2):252-65.

Brady, H. E. and D. Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (2nd Ed.). Lanham: Rowman & Littlefield Publishers.

Bunea, A. and F. R. Baumgartner. 2014. "The State of the Discipline: Authorship, Research Designs, and Citation Patterns in Studies of Eu Interest Groups and Lobbying." *Journal of European Public Policy* 21(10):1412-34.

Collier, D. 1993. "The Comparative Method." Pp. 105-19 in *Political Science: The State of the Discipline Ii*, edited by Ada Finifter. Washington D.C.: American Political Science Association.

Collier, D., H. E. Brady, and J. Seawright. 2004. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." Pp. 229-66 in *Rethinking Social Inquiry. Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier. Lanham: Rowman & Littlefield.

Coppedge, M. and D. Kuehn. 2019. "Introduction: Absorbing the Four Methodological Disruptions in Democratization Research?" *Democratization* 26(1):1-20.

Elman, C. 2013. "Duck-Rabbits in Social Analysis: A Tale of Two Cultures." *Comparative Political Studies* 46(2):266-77.

George, A. L. and A. Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, Mass.: MIT Press.

George, A. L. and T. J. McKeown. 1985. "Case Studies and Theories of Organizational Decision Making." Pp. 21-58 in *Advances in Information Processing in Organizations*, Vol. 2, edited by Robert F. Coulam and Richard A. Smith. Greenwich, CT: JAI Press.

Gerring, J. 2017. "Qualitative Methods." *Annual Review of Political Science* 20(1):15-36.

Giles, M. W. and J. C. Garand. 2007. "Ranking Political Science Journals: Reputational and Citational Approaches." *PS: Political Science & Politics* 40-(4):741-51.

Goertz, G. and J. Mahoney. 2012. *A Tale of Two Cultures: Contrasting Qualitative and Quantitative Paradigms*. Princeton: Princeton University Press.

Goertz, G. and J. Mahoney. 2013a. "For Methodological Pluralism: A Reply to Brady and Elman." *Comparative Political Studies* 46(2):278-85.

Goertz, G. and J. Mahoney. 2013b. "The Third Wave of Qualitative Methodology." *Newsletter of the APSA Section on Qualitative and Multi-Method Research* 11(1):14-8.

Groeneveld, S., L. Tummers, B. Bronkhorst, T. Ashikali, and S. van Thiel. 2014. "Quantitative Methods in Public Administration: Their Use and Development Through Time." *International Public Management Journal* 18(1):61-86.

Hall, P. A. 2008. "Systematic Process Analysis: When and How to Use It." *European Political Science* 7(3):304-17.

Hsueh, R. 2012. "China and India in the Age of Globalization: Sectoral Variation in Postliberalization Reregulation." *Comparative Political Studies* 45(1):32-61.

Hopf, T. 1991. "Polarity, the Offense-Defense Balance, and War." *American Political Science Review* 85(2):475-493.

Hu, L.-T. and P. M. Bentler. 1995. "Evaluating Model Fit." Pp. 76-99 in *Structural Equation Modeling. Concepts, Issues, and Applications*, edited by Rick H. Hoyle. Thousand Oaks: Sage.

Keele, L. 2015. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis* 23(3):313-35.

King, G., R. O. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Koivu, K. L. and E. K. Damman. 2015. "Qualitative Variations: The Sources of Divergent Qualitative Methodological Approaches." *Quality & Quantity* 49-(6):2617-32.

Kuehn, D. and I. Rohlfing. 2016. "Are There Really Two Cultures? A Pilot Study on the Application of Qualitative and Quantitative Methods in Political Science." *European Journal of Political Research* 55(4):885-905.

Lebo, M. J. and C. Weber. 2015. "An Effective Approach to the Repeated Cross-Sectional Design." *American Journal of Political Science* 59(1):242-58.

Lijphart, A. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65(3):682-93.

Mahoney, J. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62(1):120-47.

Mahoney, J. and G. Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14(3):227-49.

Marx, A., B. Rihoux, and C. Ragin. 2014. "The Origins, Development, and Application of Qualitative Comparative Analysis: The First 25 Years." *European Political Science Review* 6(1):115-42.

Moses, J., B. Rihoux, and B. Kittel. 2005. "Mapping Political Methodology: Reflections on a European Perspective." *European Political Science* 4: 55-68.

Munck, G. L. and R. Snyder. 2007. "Debating the Direction of Comparative Politics: An Analysis of Leading Journals." *Comparative Political Studies* 40(1):5-31.

Pepinsky, Thomas B. 2019. "The Return of the Single-Country Study" *Annual Review of Political Science* 22: 187-203.

Rohlfing, I. 2013. "A Tale of Four Books, Or: Principles and Practice of Social Science Methods." *Newsletter of the APSA Section on Qualitative and Multi-Method Research* 11(1):12-4.

Schneider, C. Q. and C. Wagemann. 2013. "Are We All Set?" *Newsletter of the APSA Section on Qualitative and Multi-Method Research* 11(1):5-8.

Schwemmer, C. and O. Wieczorek. 2020. "The Methodological Divide of Sociology - Evidence from Two Decades of Journal Publications." *Sociology*. 54(1):3-21.

Traxler, F. 1992. "Interests, Politics, and European Integration." *European Journal of Political Research* 22(2):193-217.

Weyland, K. 2010. "The Diffusion of Regime Contention in European Democratization, 1830-1940." *Comparative Political Studies* 43(8):1148-1176.