Paired Comparison and Theory Development: Considerations for Case Selection

Rachel M. Gisselquist, *United Nations University*

ABSTRACT Despite the widespread use of paired comparisons, we lack clear guidance about how to use this research strategy in practice, particularly in case selection. The literature tends to assume that cases are systematically selected from a known population, a major assumption for many topics of interest to political scientists. This article speaks to this gap. It describes three distinct logics of paired comparison relevant to theory development, presents a simple way of considering and comparing them, and explores how this approach can inform more intentional research design, with particular attention to low information settings where substantial research is needed to ascertain the values of independent or dependent variables. The discussion underscores inter alia the need to be aware and explicit about the implications of case selection for the ability to test and build theory, and the need to reconsider the well-cited "rule" of not selecting on the dependent variable.

Paired comparisons have long been a staple of research in political science. Comparative politics, in particular, is commonly defined by the comparative method, one approach to paired comparison (see Lijphart 1971, 682; 1975, 163; Slater and Ziblatt 2013). Despite their widespread use, however, we lack a "theory of practice" (Tarrow 2010). The methodological literature has not paid much attention to paired comparisons, focusing instead on single case studies and large-n analysis. King, Keohane, and Verba's classic *Designing Social Inquiry,* for instance, offers no specific strategies distinct from those of other forms of multicase qualitative analysis (King, Keohane, and Verba 1994, 43–46). This lacuna in the literature poses particular challenges for scholars using paired comparative methods, especially in their initial selection of cases. To the extent that the literature has addressed case selection for paired comparisons, it has assumed that researchers know ex-ante about the universe of cases and can select cases systematically from a known population (see George 1979; Plümper, Troeger, and Neumayer n.d.; Seawright and Gerring 2008, 296). This is a big assumption for many topics of interest to political scientists, including ethnic conflict, new political parties, protests, and social movements, all of which are "almost impossible to observe and to clearly delimit" (Hug 2003, 257). More broadly, a common complaint among political scientists is that the case selection guidelines that the literature does suggest are, in practice, impossibly difficult to follow. Indeed, it is quite common for scholars to admit that their cases were selected for largely practical reasons and paired comparisons were constructed post hoc, with finesse.

Building on the literature, and especially on Tarrow (2010)'s effort to move us toward a theory of practice, this article speaks to this gap. It argues that there are three distinct logics of paired comparison for theory development, presents a straightforward way of considering and comparing them using simple tables, and explores how this approach can be used to inform more intentional research design using paired comparisons. Low information settings, in which substantial research is needed to ascertain the values of independent variables (IV) and dependent variables (DV), present special challenges for case selection. The article focuses on how the approach outlined here may be useful in developing research designs in such settings. In reviewing several different scenarios, the discussion suggests that there are no easy answers. It underscores inter alia the need for scholars to be aware and explicit about the implications of their case selection for their ability to test and build theory, and the need to add major qualifications to the well-cited "rule" of not selecting on the DV: a researcher interested in testing hypotheses about necessary conditions often *should* select on the DV, if information allows, and use the logic of the method of agreement to test hypotheses.

First, the article explores the use of case studies in theory development. Then, it turns to the logics of paired comparison and their implications for theory building. It concludes by exploring

Rachel M. Gisselquist *is a research fellow with the United Nations University, World Institute for Development Economics Research (UNU-WIDER). Her research focuses on the comparative politics of the developing world, with particular attention to ethnic politics, governance, and state fragility in sub-Saharan Africa. She holds a PhD in political science from MIT. She can be reached at rachel.gisselquist@gmail.com.*

how this framework can inform more intentional strategies of case selection in situations of incomplete information about IVs and DVs.

CASE ANALYSIS AND THEORY DEVELOPMENT

Paired comparisons as reviewed here involve the study of two cases selected for theory development, that is, of generating and/or testing theoretical propositions that offer generalizable causal explanations. A "case" is understood broadly in Lijphart (1975, 160)'s terms as "an entity on which only one basic observation is made and in which the independent and dependent variables do not change during the period of observation which may cover a long term, even several years."[1] This wording is somewhat confusing when the DV itself implies change, as in much comparative research (e.g., the emergence of ethnic political parties, a rise in social protest, the reconstruction of postconflict states). In such cases, the dependent variable can be understood as "change" or "lack of change" for a period of study, or in a finer-grained way with reference to characteristics of change, such as its degree or speed.

Paired comparisons differ from single case studies primarily in their ability to control for variables (see Lijphart 1971, 684; Tarrow 2010, 244). Like single-case studies, they also allow for causal-process analysis, which is generally not possible in large-n studies (Brady and Collier 2004). They can further provide a useful stepping stone from single-case to multi-case comparison (Becker 1968). Comparative analysis of more than two cases theoretically can allow for even more analytical control, but paired comparisons may have benefits over comparative analysis of more than two cases because they allow, in practice, for greater descriptive depth and more control of relevant variables (Slater and Ziblatt 2013).

Because single-case studies and paired comparisons are so closely related, much of the literature on case studies and theory development is relevant to paired comparisons (Eckstein 1975; George 1979; Gerring 2004; Lijphart 1971). Eckstein (1975) and Lijphart (1971)'s classic frameworks highlight four types of case studies relevant to theory-building and, sometimes, theory-testing: (1) "heuristic" or "hypothesis generating" case studies; (2) "plausibility probes" to assess the potential plausibility of hypotheses; (3) "theory infirming" or "theory confirming" cases, which, sometimes may be "crucial" tests of hypotheses; and (4) "deviant" cases that are not well-explained by existing theories and strategically selected for study to help refine theories. In Lijphart (1971, 692)'s view, theories should not necessarily be accepted or rejected on the basis of a single theory-confirming or theory-infirming case, but rather the cases should be seen as strengthening or weakening the hypothesis in question: if "the proposition is solidly based on a large number of cases," another theory-confirming case obviously "does not strengthen it a great deal." (Eckstein and Lijphart argue that "crucial" theory-infirming or theory-confirming cases can provide definitive tests for hypotheses, a point on which there is some debate that is beyond the scope of this article (see Gerring 2007).) Heuristic or hypothesis-generating cases are most relevant to areas of study where theory is not well-developed, while theory-infirming, theory-confirming, and deviant cases can only be conducted when developed theory offers precise, testable hypotheses.

Two other types of case studies are also common in the literature, but they are not designed for theory building, although they can contribute indirectly to it (Lijphart 1971, 691-692). These are Eckstein's "configurative-ideographic" and "disciplined-configurative" cases, both of which are primarily descriptive. The latter explicitly builds

on theoretical propositions in interpreting a case, but it does not use the case to examine the propositions themselves.

A typology of paired comparisons is analogous (see George 1979). If they conform to one of the first four types outlined previously, comparisons of two cases can arguably offer more leverage on theory development than single cases, but not all comparisons of two cases are designed to contribute directly to theory building. Many studies that explicitly use "paired" comparisons are of the configurative-ideographic and disciplined-configurative type. The rest of this article focuses on paired comparisons for theory development only.

THE LOGICS OF PAIRED COMPARISON

In his *System of Logic*, Mill described four methods of "experimental inquiry," the first two, the methods of agreement and difference, constitute the core logics of paired comparisons for the development of causal theories (Mill 1882). To these two, we add Przeworski and Teune (1970)'s "most different systems" analysis, an important variation on the method of agreement.

A large literature already exists on comparative methods, which includes detailed examples (e.g., Collier 1993; Geddes 1990; George 1979; Przeworski and Teune 1970; Skocpol 1979; Slater and Ziblatt 2013; Tarrow 2010; Van Evera 1997). Without rehashing those discussions here, this article builds on them to focus on describing the three core logics of paired comparison for theory development, using a simple framework, with figures. This framework facilitates more intentional consideration of research design options, as explored in the final section of this article. Note that in defending the method of agreement as a legitimate approach to paired comparison for theory development, this discussion shares the interpretation of, for instance, Tarrow (2010), but differs with Lijphart (1975) and King, Keohane, and Verba (1994). In highlighting the similarities between most different systems analysis and the method of agreement, it shares the interpretation of, for instance, Gerring (2007), but differs with Przeworski and Teune (1970) and Meckstroth (1975).

For simplicity, the term "variable" is used to describe the factors that may be considered in the design of paired comparisons (e.g., $w, x, y,$ and z in the figures). However, these factors can also be understood as combinations of variables, or as discrete processes or mechanisms; indeed, some of the best paired comparisons use the logics described here to conduct causal-process analysis (see Brady and Collier 2004).

Comparable Cases and the Method of Difference

The most accepted method of paired comparison is the comparable case strategy (see Abadie, Diamond, and Hainmueller 2012; Lijphart 1975). The logic of comparable cases is summarized by Mill (1882, 483) in terms of the method of difference: "If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance

Figure 1

Comparable Cases (Mill's Method of Difference)

	Independent Variables				Dependent Variable
	w	*x*	*y*	*z*	
Case I	-	-	+	+	□
Case II	+	-	+	+	■

in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon." Figure 1 illustrates with a simple example in which four IVs (w, x, y, and z) and the DV are all binary values. The IVs are either present ("+") or absent ("-"). The DV is either positive ("■") or negative ("□"). As shown, all IVs, except the circled causal variable, w, share the same value, and w varies with the DV.

More generally, comparable case analysis may match more subtle variations in IVs and DVs, which are not binary values. Strictly speaking, such analysis follows Mill's "method of concomitant variations," but the core logic is that of the method of difference for as Mill (1882, 496) notes, "the concomitance itself must be proved by the Method of Difference" (see Lijphart 1971, 688).

Przeworski and Teune (1970) discuss this method in terms of the "most similar systems" approach, highlighting its use at the "system" level, by which they generally mean the country level. In their view, because there are so many differences across countries, even when carefully selected as comparable cases, outcomes are over-determined and "experimental variables cannot be singled out" using this approach (Przeworski and Teune 1970, 34). (This criticism is related to a more general criticism of paired comparisons, insufficient degrees of freedom, which is discussed more fully in the text that follows.) Although the method of difference is often used at the national level consistent with a most similar systems approach, it need not be. The comparable case strategy can be used, for instance, with reference to diachronic comparisons within a single country, comparison of subnational units in the same country, and comparison of subnational units in several countries.

King, Keohane, and Verba (1994, 129) argue that "When observations are selected on the basis of a particular value of the dependent variable, nothing whatsoever can be learned about the causes of the dependent variable without taking into account other instances when the dependent variable takes on other values." Geddes (1990, 149) elaborates:

> This is not to say that studies of cases selected on the dependent variable have no place in comparative politics. They are ideal for digging into the details of how phenomena come about and for developing insights. They identify plausible causal variables. They bring to light anomalies that current theories cannot accommodate. In so doing, they contribute to building and revising theories. *By themselves, however, they cannot test the theories they propose and, hence, cannot contribute to the accumulation of theoretical knowledge...* [italics mine]

Many other methodologists, however, disagree—the position taken here (see, e.g., Brady and Collier 2004; Collier 1993; Van Evera 1997). Indeed, several of the ways in which comparative analysis based on the method of agreement can contribute to theoretical knowledge are precisely noted in Geddes (1990), which arguably adopts too narrow a view of theory development. In light of Eckstein (1975)'s and Lijphart (1971)'s typologies of cases, at least three contributions can be identified:

First, comparisons selected on the DV and paired on the basis of the method of agreement can identify causal variables and build theories, in the manner of hypothesis-building cases or plausibility probes. Second, they can "bring to light anomalies that current

Although the method of difference is often used at the national level consistent with a most similar systems approach, it need not be. The comparable case strategy can be used, for instance, with reference to diachronic comparisons within a single country, comparison of subnational units in the same country, and comparison of subnational units in several countries.

Method of Agreement

The method of agreement is the converse of the method of difference. Mill (1882, 482) summarizes its logic as follows: "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon." Figure 2 offers an example using the same IVs and DVs as figure 1. As shown, all IVs, except w and the DV have different values, suggesting the presence of w as the cause of the DV.

Many analysts reject the use of the method of agreement in theory development because it involves selecting on "extreme" values of the DV, thus introducing selection bias (see Collier 1995, 464).

theories cannot accommodate" and use them to refine theories, in the manner of deviant cases (Geddes 1990, 149).

Third, they can be used as theory-infirming cases with reference to theories about necessary conditions (Braumoeller and Goertz 2000; Collier 1993; Dion 1998; Jervis 1989). This latter view is consistent with Mill's interpretation that the method could not demonstrate causation, but could help to eliminate causal factors (see Collier 1993, 464). In thinking about testing theories of necessary conditions, Dion (1998, 141) offers the useful example of how to test the hypothesis that state crisis is a necessary condition for social revolution, which would involve obtaining a list of all social revolutions, selecting a random sample, and identifying cases preceded by state crisis. "One would not gather a list of state crises and then see whether they resulted in social revolutions or obtain a biased selection of the social revolutions" (Dion 1998, 133). As the example suggests, the method of agreement is most powerful for theory testing when multiple cases are examined; two cases may not provide enough information, but they could. For instance, the method could infirm the hypothesis in Dion's example if state crisis were missing from one or both of the cases of social revolution selected for analysis. (In contrast, evidence of state crisis in both cases—or even many cases, but not the entire population—would not provide

Figure 2

Mill's Method of Agreement

	Independent Variables				Dependent Variable
	w	x	y	z	
Case I	+	-	+	+	■
Case II	+	+	-	-	■

definitive confirmation of the hypothesis, but it would lend further support.)

Most Different Systems

Finally, Przeworski and Teune (1970)'s most different systems approach offers a fourth way in which the method of agreement can contribute to theory development. It hinges on exploiting different levels of analysis, which is not part of the standard method of agreement although the underlying logic is similar (Gerring 2007, 139). Figure 2 illustrates if we think of *x, y,* and *z* as system variables and *w,* the causal variable, as operating at a subsystem variable ("the level of individual actors [or]… the level of groups, local communities, social classes, or occupations") (Przeworski and Teune 1970, 34). This approach can eliminate irrelevant systemic factors in explanations of behavior at the subsystem level and thus infirm theories that operate at the systemic level (Meckstroth 1975, 137).

Effectively, most different systems analysis combines the method of agreement with multicase analysis: cases are paired at the systemic level, and multiple cases are examined at the subsystem (i.e., subnational) level. Strictly speaking, in analyzing the behavior of multiple individuals or groups at the subnational level, the approach uses the statistical method and does not suffer from the innate "few cases, small n" problem characteristic of case studies and paired comparisons (Lijphart 1975, 164).

Indeterminate Designs

Each of the three approaches to paired comparison described previously can contribute to the identification of causal variables, theory testing, and theory building (see Tarrow 2010, 235). Some comparisons of cases, however, are generally inferior for this purpose. These involve research designs in which the IVs and DVs of both cases are either both the same or both vary as shown in the examples in figure 3. On the one hand, cells A and D, which are the same as figures 1 and 2, offer examples of the methods of difference and agreement. As we have seen, assuming that we have identified all of the relevant variables—a big assumption, as discussed later—both of these methods can identify *w* as the causal variable or as a necessary condition in a causal process. The examples in cells B and C, on the other hand, are indeterminate in the sense that even if we have identified all of the relevant variables, they cannot pinpoint which of the IVs causes, or is a necessary condition for, the outcome beyond something in the set of [*x, y,* or *z*].[2]

Note, however, that cases of the sort depicted in cells B and C may infirm some specific hypotheses about necessary conditions, which could be useful in testing of some well-developed theories. For instance, in the examples here, the cases in cell B show to be false the hypothesis that the presence of *w* is always necessary for the outcome □ (it results even when *w* is absent). Similarly, cells B and C both show to be false the hypothesis that the presence of *w* precludes the outcome □ (it results both when it is present and when it is absent).

A Brief Note on Combining Comparisons

Combining paired comparisons with each other, or with multicase comparative analysis, can draw on these same logics and provide more traction on causal inference for theory development. For instance, Mill's "joint method of agreement and difference" operates as follows: "If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common save the absence of that circumstance, the circumstance in which alone the two sets of instances differ, is the effect, or the cause, of an indispensable part of the cause, of the phenomenon"

(Mill 1882, 489). Or, consider Collier and Collier (1991)'s research strategy, which begins with analysis of eight Latin American countries "roughly matched on a number of broad dimensions" and then turns to analyze "pairs of countries that are nonetheless markedly different" (Collier 1993, 112). Researchers using paired comparisons should consider such approaches building on the three core logics outlined here. However, even if the research strategy is ultimately to combine comparisons, the researcher still needs a way to think about the selection of each individual paired set, and especially of the first paired set. The rest of this article thus focuses on the selection of single paired comparisons and leaves discussion of combining comparisons to other work.

CAUTIONS AND GUIDELINES

The literature highlights various cautions and guidelines for paired comparisons. One of the core guidelines, *not selecting on the DV*, is discussed previously. Four others are well summarized by Tarrow (2010, 246-253):

- *Beware "insufficient degrees of freedom:"* The methods described previously assume all relevant variables are known and can be observed, which is a large assumption. Additional variables not controlled for in paired comparison imply insufficient degrees of freedom, that is, more uncontrolled for variables than the number of cases. This problem is inevitable, but not fatal. Our best strategy is to construct comparisons carefully, building on theory and taking advantage of new statistical methods for creative matching where possible (e.g., Seawright and Gerring 2008). However, it does imply that a paired comparison is not the best strategy if we want to test several rival hypotheses about the causal effects of specific variables against each other.

- *Beware "non-representativeness":* Selected cases may not be representative of the population, particularly if they are selected because of their extreme values. This is a common criticism, but the degree to which it is problematic is debatable. Some classic theorists consider cases useful for theory testing precisely because they have extreme values (on crucial cases, see, e.g., Eckstein 1975; Gerring 2007; Lijphart 1971).

- *Beware "atheoretical case selection":* In particular, paired comparisons are often selected "atheoretically" from one world region because of the researcher's interest in that particular region. This is inadvisable if careful attention is not also paid to research design. However, drawing comparable cases from the same region is not necessarily a bad approach as countries in the same region may share many characteristics.

- *Do not ignore scope conditions:* A paired comparison may not contribute to theory development if scope conditions are ignored in case selection. This is a problem for comparative research in general. Paired comparison may address it as progressively conducting multiple paired comparisons is one promising way of testing scope conditions (see Samuels 1999).

Several other general cautions and guidelines are suggested by other scholars. Przeworski and Teune (1970, 36), advise researchers to *use the most different systems approach and begin at the lowest level of analysis,* "most often individuals." This advice is taken to heart by some scholars (e.g., those working within a rational choice framework), but not so much by others (e.g., historical institutionalists or international relations theorists). King, Keohane, and Verba (1994) offer another guideline: *select on the IV and especially avoid*

Figure 3

Determinate and Indeterminate Paired Comparisons for Identifying Causal Variables

	DVs vary	DVs are the same
Most IVs are the same	**A.** w x y z DV case I − − + + □ case II + − + + ■	**B.** w x y z DV case I − − + + □ case II + − + + □
Most IVs differ	**C.** w x y z DV case I + − + + ■ case II + + − − □	**D.** w x y z DV case I + − + + ■ case II + + − − ■

selecting on the IV and the DV together. As King, Keohane, and Verba (1994, 143) note: "The most egregious error is to select observations in which the explanatory and dependent variables vary together in ways that are known to be consistent with the hypotheses that the research purports to test" (see also Lijphart 1975, 164). Presenting dishonest hypothesis "tests" of this sort is obviously to be avoided,

but with only two cases considered, completely ignoring the values of the IV and DV also poses major dilemmas for analysts of paired comparisons. In particular, there is a good chance of an indeterminate research design, as explored next.

CASE SELECTION IN LOW INFORMATION SETTINGS

Several studies have mapped innovative ways to select cases for comparative analysis using new statistical methods (Abadie, Diamond, and Hainmueller 2012; Seawright and Gerring 2008). These methods, however, assume significant knowledge about the population from which cases are selected. As Seawright and Gerring (2008, 296) note, "if nothing—or very little—is known about the population, the methods described in this study cannot be implemented."

What if knowledge about the population *is* incomplete? One answer is that the researcher should buckle down and compile all the necessary data on the population before proceeding (see Lieberman 2005). This can be a good strategy for some topics, but it is not a promising one for all: for many topics that we care about in political science (including ethnic conflict, new political parties, protests, and social movements), compiling a comprehensive database can be an impossible task, with the results inherently suffering from problems of selection bias (Hug 2003). Such data collection thus may involve considerable time and effort only to result in a problematic inventory. Challenges may be compounded for some countries or regions, such as much of sub-Saharan Africa, where basic data and sources such as newspaper archives may be incomplete or take considerable work to compile (see Jerven 2013). Thus, in addressing some

research questions, the best strategy for contributing to the development of causal theories may begin with paired comparison, even if cases must be selected with an incomplete view of the population. This section explores several scenarios for case selection in low information settings in light of the logics of paired comparison mapped earlier.

WHEN THE DV IS UNKNOWN

In some situations, significant research may be needed to determine the values of the DV. If this seems unlikely for DVs with binary values like ■ and □, do not forget topics in which the DV requires more nuanced assessment. For instance, how does the quality of public institutions compare across post-civil war African countries? Or how have political parties played the ethnic card in campaign messages, including through the use of "coded" appeals (see Chandra 2009; Mendelberg 2001). Measuring the DV in such situations may be a major component of a research project itself, meaning that cases must be initially chosen with incomplete knowledge about outcomes.

> *The methods described previously assume all relevant variables are known and can be observed, which is a large assumption. Additional variables not controlled for in paired comparison imply insufficient degrees of freedom, that is, more uncontrolled for variables than the number of cases. This problem is inevitable, but not fatal.*

This would seem an ideal situation for adherents of King, Keohane, and Verba (1994): One cannot but follow their advice to select on the IV of interest and not to select on the IV and the DV together. But it also poses major risks for researchers. Consider figure 3 but imagine that all the DVs noted are unknown. In selecting cases, researchers hope to have a determinate research design (i.e., to be in cells A or D), but because they do not know the values of the DV, they may end up with an indeterminate research design (in cells B or C). King et al.'s advice points to the adoption of a strategy corresponding to the first row of figure 3, that is, to choose cases with variation in the (hypothesized) causal variable and with otherwise "similar" characteristics. If it turns out that our researchers are operating in cell A, they have a determinate research design, but they could be operating in cell B.

On the contrary, the approach introduced here suggests that cases should be selected and their selection justified—not with reference to such general guidelines—but based explicitly on the types of theoretical propositions under examination and the known characteristics of the possible cases. For instance:

Suppose our hypothesis is that the presence of *w* (e.g., national unity) is a necessary condition for the outcome ■ (e.g., successful democratic transition). Such a hypothesis might build on theories of democracy that highlight national unity as the single precondition for democratic transition (Rustow 1970). As discussed previously, in addressing hypotheses about necessary conditions, the method of agreement is better suited than the method of difference. In this instance, our hypothesis gives us a clear prediction for the DV when

w is absent (■ will not occur), but not when it is present (either ■ or □ could occur, i.e., according to our hypothesis national unity may be present in countries that do not transition to democracy, although it is also a necessary precondition for countries that do transition). Thus, it seems our preferred option here is to select two cases in which *w* is absent in both and all other IVs are the same. This would be like operating in the second row of figure 4. (Figure 4 is the same as figure 3, except that the *w*'s are absent in the second row for the purposes of simplicity in this example.) If we are "lucky," we have a research design that provides a definitive test of our hypothesis and infirms it—for example, we find, as in cells C or D in figure 4, that ■ can be the outcome even when *w* is absent. If we had instead operated in the first row of figure 4, we would have found in this particular example that cell A suggests support for our hypothesis (as ■ occurs when *w* is present and not when it is absent), while cell B does not tell us anything about our hypothesis as ■ does not occur.

On the one hand, we might consider selecting two cases in which *w* is present in both and all other IVs are the same. This would be like operating in the second row of figure 3. In this example, we find that none of the possible outcomes is a determinate research strategy: If we find that ■ is the outcome in both cases (as in cell D), our data suggest support for our hypothesis (but not a definitive test). Cell C also provides some support as ■ occurs when *w* is present (and it also shows than □ can occur when *w* is present).

Suppose, on the other hand, that we are interested in testing a theory about how *w* (e.g., modernization) "causes" ■ (e.g., democratic transition). Our theory also implies that ■ should not result when *w* is absent and that □ (i.e., failure to transition) will not result when *w* is present. Here, the standard advice to select paired cases with variation in the causal variable *w* and no variation in other IVs (i.e., to operate in the first row of figure 3) seems appropriate. If it turns out that the DV varies (as in cell A), we have a determinate

research design that also lends support to our hypothesis. If instead the DVs are the same (as in cell B), this approach would help to infirm our hypothesis—but would not allow us to identify one single causal variable from among the other IVs. Here, selecting for variation in *w* is generally a better strategy than selecting for similarity in *w*. In this particular example, if it turns out that the DVs are the same and both ■ (cell D in figure 3), our analysis lends partial support to our hypothesis, but does not speak to instances in which *w* is absent. (In contrast, in a different example in which the DVs were both □, our analysis would have helped to infirm our hypothesis.) If it turns out that DVs vary (cell C), Case II would be theory-infirming, but the cases together would not identify the causal variable.

In all of these scenarios, the possibility that ■ is a rare occurrence with little chance of occurring when we select cases with no attention to the DV, raises particular challenges for our research strategy. War is one example of a rare event (see King and Zeng 2001). In the first scenario, in particular, if the outcomes in the second row of figure 4 had all been □, we would not have learned anything new about our hypothesis. If we think it is likely that the DV of interest is a rare event, choosing cases when our incomplete knowledge suggests that it occurs would be advisable. If it turns out that we are wrong, we can still use the case studies for developing insights and perhaps in theory building, being careful to note their limits in theory testing. (Alternatively, we may also reconsider the project of compiling comprehensive information about the instances in which ■ occurs, even if the resulting database is likely to suffer from selection bias.)

When the IVs Are Unknown

In many situations, we know the values of the DV, but we have incomplete information about the explanatory and control variables. This may result either because we have identified all of the relevant variables but we do not know all of their values, or because theory is relatively weak and we are unsure of whether we have identified all of the relevant variables. Here, too, cases would need to be chosen initially with incomplete information. In this situation, the researcher is unable to follow the standard advice of selecting on the IVs. This dilemma is depicted in figure 5, which is the same as figure 3 except that cells A and B have opposite outcomes for the purposes of this example.

In this situation, the best strategy also depends on the types of theoretical propositions under examination and the particular characteristics of possible cases: suppose, as in the first instance described above, that our hypothesis is that the presence of *w* is a necessary condition for the outcome ■. Here, the best option is to select two cases in which ■ occurs (the second column of figure 5). In this particular example, our analysis will either provide support for the hypothesis (cell D), or it will infirm the hypothesis for these cases (cell B). Cell B, unfortunately, will not shed additional light for us on which of the IVs *x*, *y*, and *z* is the causal variable or necessary condition. Alternatively, selecting cases in which the DV varies gives us basically one less shot at infirming our hypothesis. In this particular example, however, Case I in cell A happens to be theory-infirming

Figure 4
Testing Hypotheses about Necessary Conditions

		DVs vary						DVs are the same				
	A.	*w*	*x*	*y*	*z*	DV	B.	*w*	*x*	*y*	*z*	DV
Most IVs are the same	case I	–	–	+	+	□	case I	–	–	+	+	□
	case II	+	–	+	+	■	case II	+	–	+	+	□
	C.	*w*	*x*	*y*	*z*	DV	D.	*w*	*x*	*y*	*z*	DV
Most IVs differ	case I	–	–	+	+	■	case I	–	–	+	+	■
	case II	–	+	–	–	□	case II	–	+	–	–	■

Figure 5
When the IVs are Unknown

		Outcomes vary						Outcomes are the same				
	A.	*w*	*x*	*y*	*z*	DV	B.	*w*	*x*	*y*	*z*	DV
Similar characteristics	case I	?(-)	?(-)	+	+	■	case I	?(-)	?(-)	+	+	■
	case II	?(+)	?(-)	+	+	□	case II	?(+)	?(-)	+	+	■
	C.	*w*	*x*	*y*	*z*	DV	D.	*w*	*x*	*y*	*z*	DV
Different characteristics	case I	?(+)	?(-)	+	+	■	case I	?(+)	?(-)	+	+	■
	case II	?(+)	?(+)	–	–	□	case II	?(+)	?(+)	–	–	■

as ■ occurs even if *w* is absent, while cell C suggests some partial support for the hypothesis.

Suppose then, as in the second situation discussed earlier, that we are interested in testing a theory about how *w* "causes" ■. Here it is advisable to select paired cases with variation on the DV (i.e., the first column of figure 5). In this particular example, cell A suggests exactly the opposite relationship to our hypothesis and is theory-infirming. Cell C is also inconsistent with it. If we had instead chosen for similar DVs, cell B would also have been inconsistent with our hypothesis (but provided no additional traction on the causal variable), while cell D would have suggested some support.

More broadly, this article suggests the need for appropriate modesty in describing what can be learned with reference to theory from a single paired comparison, however elegantly it is constructed. Both the challenges of constructing paired comparisons well, and the limits to a number of comparative strategies are illustrated. Indeed, particularly in low information settings, the discussion suggests that it is the rare research project that will uncover a paired comparison that can provide a definitive test of a theory. Indeed, without a full view of the population from which the cases are selected in many instances, it can be unclear even whether these cases are "extreme" or "representative." However, when compared to other options like building

> *More broadly, this article suggests the need for appropriate modesty in describing what can be learned with reference to theory from a single paired comparison, however elegantly it is constructed.*

Incomplete Information about Both IVs and DVs

In reality, at the start of their work researchers often face a situation that looks more like figure 6, with some data but overall incomplete information about both IVs and DVs. In such situations, the advisable strategy, still seems to depend on the type of hypotheses one is testing and the available data. However, particularly when information is this incomplete, researchers should be prepared to temper their claims and ambitions of theory testing through paired comparison, unless they luckily happen on a design that can definitively test their hypotheses.

CONCLUSION

Building on the literature, and especially on Tarrow (2010)'s points toward a theory of practice, this article identifies three logics underlying the use of paired comparisons for theory development and presents a simple approach, with visualizations, that might help researchers to build on these logics to develop and defend more intentional research designs using paired comparisons. The final section of the article illustrates how this approach can be used to explore the implications for case selection of various scenarios in low information settings, where knowledge about IVs and/or DVs is incomplete. Such situations, are relatively common for political scientists addressing core questions in the field. This discussion suggests no easy solutions. In general, it underscores, among other points, the need to add a number of qualifications to the "rule" of not selecting on the DV.

a large-n dataset, paired comparison may still provide the best strategy for theory development on many low information topics.

Finally, as briefly noted, a paired comparison is a problematic approach in terms of degrees of freedom if the researcher seeks to test multiple theories against each other. In other words, implicit in this discussion is the notion that paired comparative research will be guided primarily by focus on a single working theory. However, combined with each other or other data, paired comparison can be a core component of a research strategy designed to test multiple theories. For instance, multiple paired comparisons might be used together to triangulate the set of causal variables that explain a particular phenomenon.

NOTES

1. For further discussion about the definition of a case see, e.g., Gerring (2004); Ragin and Becker (1992).

2. A critical reader might note that the outcome could be caused by more than one variable, e.g., in cell C, the outcome could be caused by a process involving *x*, *y*, and *z*. This criticism is somewhat misleading: In the example w, *x*, *y*, and *z* represent all of the ways in which cases I and II might meaningfully be compared. It is unlikely that "almost everything" explains the DV. In addition, as noted above, *w*, *x*, *y*, and *z* are described as variables purely for descriptive simplicity; they might equally well be understood as combinations of multiple variables or processes themselves.

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2012. "Comparative Politics and the Synthetic Control Method." In. Unpublished paper available at http://www.hks.harvard.edu/fs/aabadie/comparative.pdf (last accessed 28 January 2013).

Figure 6
Incomplete Information about IVs and DVs

			(Outcomes seem to vary)						(Outcomes seem to be the same)			
	A.	*w*	*x*	*y*	*?*	DV	**B.**	*w*	*x*	*y*	*?*	DV
(Potential causal variable seems to vary)	case I	-	-	-	?	■	case I	-	-	-	?	□
	case II	+?	-	-	?	□?	case II	+?	-	-	?	□
	C.	*w*	*x*	*y*	*?*	DV	**D.**	*w*	*x*	*y*	*?*	DV
(Potential causal variable seems to be the same)	case I	+	+	?	?	■	case I	+	?	+	?	□
	case II	+?	-	-	?	□	case II	+?	?	-	?	□?

Becker, Howard S. 1968. "Social Observation and Social Case Studies." In *International Encyclopedia of the Social Sciences,* 232–238. New York: Macmillan.

Brady, Henry E., and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, MD: Rowman and Littlefield.

Braumoeller, Bear F., and Gary Goertz. 2000. The Methodology of Necessary Conditions. *American Journal of Political Science* 44 (4): 844–858.

Chandra, Kanchan. 2009. "A Constructivist Dataset on Ethnicity and Institutions." In *Measuring Identity: A Guide for Social Scientists*, eds. Rawi Abdelal, Yoshiko M. Herrera, Alastair Iain Johnston and Rose McDermott. Cambridge: Cambridge University Press. 250–278.

Collier, David. 1993. "The Comparative Method." In *Political Science: The State of the Discipline II*, ed. Ada W. Finifter, 105–119. Washington, DC: American Political Science Association.

Collier, Ruth Berins, and David Collier. 1991. *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America.* Princeton, NJ: Princeton University Press.

Dion, Douglas. 1998. Evidence and Inference in the Comparative Case Study. *Comparative Politics* 30 (2): 127–145.

Eckstein, Harry. 1975. "Case Study and Theory in Political Science." In *Handbook of Political Science*, eds. F.I. Greenstein and N.W. Polsby, 79–138. Reading, MA: Addison-Wesley.

Geddes, Barbara. 1990. How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics. *Political Analysis* 2 (1): 131–150.

George, Alexander. 1979. "Case Studies and Theory Development: The Method of Structured, Focused Comparison." In *Diplomacy: New Approaches in History, Theory, and Policy*, ed. Paul Gorden Lauren. New York: Free Press. 43–68.

Gerring, John. 2004. What Is a Case Study and What Is It Good for? *American Political Science Review* 98 (02): 341–354

———. 2007. *Case Study Research: Principles and Practices.* New York: Cambridge University Press.

———. 2007. Is There a (Viable) Crucial-Case Method? *Comparative Political Studies* 40 (3): 231-253.

Hug, Simon. 2003. Selection Bias in Comparative Research: The Case of Incomplete Data Sets. *Political Analysis* 11 (3): 255–274.

Jerven, Morten. 2013. *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It.* Ithaca, NY: Cornell University Press.

Jervis, Robert. 1989. Rational Deterrence: Theory and Evidence. *World Politics* 41 (2): 183–207.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton, NJ: Princeton University Press.

King, Gary, and Langche Zeng. 2001. Explaining Rare Events in International Relations. *International Organization* 55 (3): 693–715.

Lieberman, Evan S. 2005. Nested Analysis as a Mixed-Method Strategy for Comparative Research. *American Political Science Review* 99 (03): 435–52.

Lijphart, Arend. 1971. Comparative Politics and the Comparative Method. *The American Political Science Review* 65 (3): 682–693.

———. 1975. II. The Comparable-Cases Strategy in Comparative Research. *Comparative Political Studies* 8 (2): 158–77.

Meckstroth, Theodore W. 1975. "Most Different Systems" and "Most Similar Systems:" A Study in the Logic of Comparative Inquirty. *Comparative Political Studies* 8 (2): 132–57.

Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality.* Princeton, NJ: Princeton University Press.

Mill, John Stuart. 1882. *A System of Logic, Ratiocinative and Inductive.* 8th. Republished as Project Gutenberg Ebook, released 31 January 2009. ed. New York: Harper & Brothers.

Plümper, Thomas, Vera E. Troeger, and Eric Neumayer. n.d. "Case Selection and Causal Inference in Qualitative Research."

Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry.* New York: Wiley-Interscience.

Ragin, Charles C., and Howard Saul Becker, eds. 1992. *What Is a Case?: Exploring the Foundations of Social Inquiry.* Cambridge: Cambridge University Press.

Rustow, Dankwart A. 1970. Transitions to Democracy: Toward a Dynamic Model. *Comparative Politics* 2 (3): 337–64.

Samuels, Richard J. 1999. Nudging Serendipity: The 3-C Problem in Comparative Politics. *APSA/CP, Newsletter of the Comparative Politics Section of the APSA* 10: 13–16.

Seawright, Jason, and John Gerring. 2008. Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. *Political Research Quarterly* 61 (2): 294–308.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia and China.* New York: Cambridge University Press.

Slater, Dan, and Daniel Ziblatt. 2013. The Enduring Indispensability of the Controlled Comparison. *Comparative Political Studies* 46 (10): 1301–1327.

Tarrow, Sidney. 2010. The Strategy of Paired Comparison: Toward a Theory of Practice. *Comparative Political Studies* 43 (2): 230–59.

Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science.* Ithaca, NY: Cornell University Press.