The Methodology of Necessary Conditions

Bear F. Braumoeller Harvard University
Gary Goertz University of Arizona

Necessary conditions provide an interesting example of a concept that everyone knows, that many people use, and yet for which no explicit methodology exists. The gap between theory and empirical testing in political science is rarely as wide as it is in the case of necessary conditions. Political science is rich in theories and hypotheses that imply necessity, but adequate empirical tests are lacking. As the concept of a necessary condition is a useful one for social scientists, methodological tools for the evaluation of necessary condition hypotheses must be developed. This constitutes our purpose. We describe appropriate procedures for the two key aspects of the empirical evaluation of necessary condition hypotheses: determining (1) whether *X* is a necessary condition for *Y* and, if so, (2) whether *X* is trivially necessary.

The idea of a necessary condition is so pervasive that the profession does not even consider it to be a concept that needs to be taught or discussed: virtually no methodology class covers it, nor does it exist as a topic in methodology texts (for an exception, see Most and Starr [1989]). In some ways these facts are not surprising, since the concept of a necessary condition seems quite basic. Nevertheless, we contend that the use of necessary conditions in social science research deserves attention because (1) necessary conditions have specific theoretical properties that are not widely understood and that therefore merit elaboration; (2) standard statistical reflexes mislead in evaluating necessary condition hypotheses; and (3) despite the discipline's collective methodological sophistication, no tests of necessary condition hypotheses that have been utilized to date are both correct and complete.

Why Do We Need a Methodology of Necessary Conditions?

To someone who does not believe in the existence of necessary conditions, a methodology of necessary conditions may seem irrelevant. If it were known with certainty that no necessary conditions exist, that would be true—*but it is impossible to obtain that knowledge without conducting tests of necessary condition hypotheses.* The need for any methodology is based on the number and importance of the theories that require it, not on *a priori* beliefs about the causal structure of the universe.

Granted, it may be difficult to think of important examples of necessary condition hypotheses in one's own area of study, and the first examples that come to mind may seem trivial:

Bear F. Braumoeller is Assistant Professor of Government, Weatherhead Center for International Affairs, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138. Gary Goertz is Assistant Professor of Political Science, University of Arizona, Tucson, AZ 85721 (ggoertz@u.arizona.edu).

The order of the authors' names is arbitrary. This research was supported, in part, by a National Science Foundation Grant SES–9309840. We would especially like to thank Doug Dion for getting us to think about empirical tests of necessary condition hypotheses. We also thank Chris Achen, Nancy Burns, James Morrow, Bob Pahre, David Rousseau, Bruce Russett, Brian Sala, Anne Sartori, and Harvey Starr for helpful comments and Bruce Bueno de Mesquita and Bruce Russett for making their data available to us for analysis.

American Journal of Political Science, Vol. 44, No. 4, October 2000, Pp. 844–858

TABLE 1 Necessary Condition Hypotheses Tested in This Article

ANDERSON AND MCKEOWN 1987: For a war to occur between two nations, this conception implies the following necessary conditions: 1. At least one of the nations is experiencing a disparity between achievements and aspirations. 2. There is a history of previous interaction that leads this nation to focus attention on the other as a possible target for military action. 3. Rule-of-thumb calculations convince at least one set of leaders that going to war has a reasonable chance of producing an acceptable outcome.
BUENO DE MESQUITA 1981: The results just reported strongly support the proposition that positive expected utility is necessary—though not sufficient—for a leader to initiate a serious international dispute, including a war.
BUENO DE MESQUITA AND LALMAN 1992: (Realpolitik Proposition 3.1: The Negotiation/Status Quo Theorem.) Under full information conditions and with demands being exogenous to the international interaction game, only negotiation or the status quo can be an equilibrium outcome of the realpolitik variant of the game.
DIAMOND 1992: It is important to emphasize as well that democracy can occur at low levels of development if the crucial mediating variables are present. Economic development is not a prerequisite of democracy.
GOERTZ AND DIEHL 1995: Not every political shock will produce a new rivalry or end existing ones. There are many conditions (not all known or specified here) for the start or conclusion of a rivalry and a political shock is only one of them. Thus, some political shock is a necessary condition for the initiation and termination of an enduring rivalry. But it is not a sufficient condition . . .
KUGLER AND ORGANSKI 1989: Clearly, the necessary but not sufficient conditions for major war emerge only in the rare instances when power parity is accompanied by a challenger overtaking a dominant nation.
OSHERENKO AND YOUNG 1993: Our examination . . . suggests to us that leadership exercised by individuals is a necessary condition for regime formation.
OSTROM 1991: By 'design principle' I mean an essential element or condition that helps to account for the success of these institutions . . . I am willing to speculate . . . [that] it will be possible to identify a set of necessary design principles and that such a set will contain the core of what has been identified here.
ROSENTHAL 1969: Inclusion of the Modérés in any R.P.F. alliance may have been necessary for the R.P.F. to retain the support of conservative voters. . . . Only 2 of the 13 R.P.F. coalitions occurred in districts without Modéré incumbents, suggesting that the presence of such an incumbent can be regarded as a necessary condition for an R.P.F. alliance.
RUSSETT 1995: Neither . . . nor shared democracy is a necessary condition for avoiding war. But, allowing for some possibility of irrationality or misconception, either may well constitute a virtually sufficient condition.
RYAN 1995: USTR [US Trade Representative] sometimes initiated 301 investigations if one of the two conditions [high commercial competitiveness or high GATT regime utility] was met, but never initiated 301 investigations if neither of the conditions was met.
SCHWELLER 1992: 1. A power transition involving a declining democratic leader is both a necessary and sufficient condition for the absence of preventive war. 2. A power transition involving a declining nondemocratic leader is a necessary but not sufficient condition for a preventive war, regardless of the regime type of the challenger.
SIVERSON AND STARR 1989: Whether or not it is the border/opportunity or willingness/alliance variables which are best able to account for the war behavior of states, it is clear that both of them together produce some powerful underlying necessary conditions for states to be at war.

> The search for necessary conditions is problematic because the utility of a necessary condition is poorly understood. There are an infinite number of necessary conditions for any phenomenon. For example, it is true that all armies require water and gravity to operate, but the contribution of such universals is modest. (Downs 1989, 234)

Concluding that necessary condition hypotheses are either rare or trivial or both, however, would be unwarranted. The necessary condition hypotheses that we test herein, listed in Table 1, cover a wide range of substantively important topics, including environmental regimes, enduring rivalries, collective action, American politics, French politics, democratization, cognitive psychology, international political economy, and war. Moreover, as we will demonstrate, none of the hypotheses that pass our test is trivial in the sense implied above.

The Concept of a Necessary Condition

We now consider the concept of a necessary condition and its empirical implications. For simplicity, we initially assume a dichotomous X and Y. In terms of Y we use the terms "occurrence" or "success" (Y) and "non-occurrence" or "failure" ($\sim Y$). For X we use the terms "present" (X) or "absent" ($\sim X$).[1] Necessity implies that Y occurs *only if* X is present. Two empirical statements logically follow. The first is that $P(X|Y) = 1$:

> Definition 1: X is a necessary condition for Y if X is always present when Y occurs.

If X is necessary for Y then the absence of X is sufficient for the absence of Y;[2] therefore, another formulation of a necessary condition hypothesis is $P(Y|\sim X) = 0$.

> Definition 2: X is a necessary condition for Y if Y does not occur in the absence of X.

Testing the hypothesis that X is necessary for Y requires translating our definitions into empirical propositions. If we represent X and Y as dichotomous variables with absence and presence coded as 0 and 1, respectively, this can be done in two ways, as Figure 1 demonstrates. In terms of the table, $P(X|Y)$ becomes $P(X = 1|Y = 1)$; Proposition 1 posits that $P(X = 1|Y = 1) = 1$. When testing this proposition, cases in which $Y = 0$ *are irrelevant to ascertaining the necessity of X for Y* and need not be sampled. We emphasize this point in Figure 1 by putting dashes in the $Y = 0$ cells. By contrast, Proposition 2 posits that $P(Y = 1|X = 0) = 0$. Here, analysis focuses on cases in which X does *not* occur; hence, when testing this proposition, $X = 1$ cases are irrelevant.

Both propositions follow from necessity, but each implies a different testing strategy. The democratic peace literature provides an excellent illustration of this point. Most democratic peace scholars collect a sample of wars and then measure the democracy of the participants (Proposition 1); on the other hand, Doyle (1983) collected a sample of liberal states and noted that none has gone to war with another (Proposition 2). Since the propositions are theoretically equivalent the choice in practice is likely to be determined by practical data gathering issues, such as the relative ease and cost-effectiveness of finding cases of complete certainty versus finding cases of war, or by coding rules (a "non-war", for example, is much more nebulous than a war, which makes a test of Proposition 1 more meaningful.) The data examined herein lend themselves more readily to tests of Proposition 1, so we will utilize it exclusively, with the understanding that tests of Proposition 2 are no less plausible.

FIGURE 1 *X* Is Necessary for *Y*

Proposition 1

Y \ X	0	1
0	—	—
1	0	100

Proposition 2

Y \ X	0	1
0	100	—
1	0	—

Scholars using quantitative methods typically fall victim to their natural tendency to utilize data from all four cells of the table when attempting to evaluate necessary condition hypotheses. Doing so when evaluating hypotheses of necessity is inherently wrong. Because cases in which $X = 1$ and $Y = 0$ are irrelevant to the hypothesis, tests based on all values of X and Y are influenced by irrelevant data and therefore provide incorrect results.

If anything, misunderstanding the implications of necessary conditions is even more dangerous for scholars using qualitative methodology. Whereas quantitative studies can (as we demonstrate below) be reexamined if only a subset of the data is relevant, case study researchers who fail to appreciate the logic of necessity typically gather too little data rather than too much. For example, Lebow (1981) tested four hypothesized necessary conditions for deterrence success by examining thirteen cases of deterrence failure and finding that the hypothesized conditions were present in eight. As Figure 2 demonstrates, however, his conclusion—that the conditions are not necessary—does not follow from the evidence, since no data were gathered in the relevant cell (deterrence successes in which the conditions are absent.) Eight of thirteen cases were examined in vain, and no reanalysis is possible.

[1] Though dichotomization may seem restrictive, even complex n-dimensional tables can often be reduced to two dichotomous dimensions. For example, Kugler and Organski (1989, 179) hypothesize that both power parity and a power transition are both necessary for great-power war; for the purposes of testing their hypothesis we need only collapse the 'no parity' and 'no transition' categories.

[2] Interestingly, any necessary condition hypothesis can be converted into a sufficient condition hypothesis and vice-versa: if $P(Y|\sim X) = 0$ (X is necessary for Y), we know that $P(\sim Y|\sim X) + P(Y|\sim X) = 1$ (that is, either Y or $\sim Y$ must happen). Because the second term on the left-hand side equals zero, we are left with $P(\sim Y|\sim X) = 1$—or, the absence of X is sufficient for the absence of Y. This point should give pause to those who, like George and Smoke (1989, 177–178), consider sufficient-condition theories to be 'much more ambitious' than necessary condition theories because of their superior predictive power and, therefore, utility.

FIGURE 2 Necessary Conditions for Deterrence Success

		Conditions absent	Conditions present
Deterrence	failure	5	8
	success	—	—

Source: Lebow (1981).

Existing Tests of Necessary Condition Hypotheses

Before we outline our method of testing necessary condition hypotheses, we must discuss techniques currently in use and their shortcomings. The only explicit statistical tests of necessity that have been used to our knowledge are Yule's Q (Bueno de Mesquita 1981; Bueno de Mesquita and Lalman 1992) and the ∇ ('del') statistic (Hildebrand, Laing, and Rosenthal 1976; Anderson and McKeown 1987; Siverson and Starr 1989), which we will argue are not measures of necessity. Other statistical techniques are occasionally utilized to evaluate theories that imply necessity as well, though the authors may be unaware of this fact. We can hardly evaluate the implications of necessity for every technique, but we can address those that have been used in order to demonstrate that existing statistical tests provide meaningless results when applied to necessary condition hypotheses.

One-way Measures of Association: Yule's Q

One statistic that has been put forth by Bueno de Mesquita (1981; Bueno de Mesquita and Lalman 1992) as a test of necessary conditions is Yule's Q, or γ. As it continues to be used despite criticism (e.g., Majeski and Sylvan 1984; Simowitz and Price 1990), it deserves brief discussion here. Our contention is that Yule's Q is of no use in testing for necessity because neither its value nor its level of significance is predicated exclusively, or even primarily, upon necessity.

As we have seen, necessity implies that $P(Y = 1|X = 0) = 0$. If this is the case, Yule's Q equals 1 (or something very close to it, if measurement error is taken into account.) That does not imply, however, that the converse is true: If Yule's Q is close to 1, X may *or may not* be necessary for Y. The simplest illustration of this fact is the case in which $P(Y = 0|X = 1) = 0$. This probability has nothing whatsoever to do with necessity—regardless of whether one chooses to test the first or second necessity proposition, the cell $X = 1/Y = 0$ is irrelevant—yet a relative scarcity of cases in this cell will cause Yule's Q to approximate 1 for any values of $P(Y = 1|X = 0)$ (for the relevant formulas see Agresti 1990). Because Yule's Q is influenced by data that are irrelevant to the hypothesis of necessity, it is not a viable test of necessary condition hypotheses.

Proportional Reduction in Error Measures: The ∇ Statistic

Another statistic used to gauge necessity is ∇, pioneered by Hildebrand, Laing, and Rosenthal (1976) and used by Anderson and McKeown (1987) and Siverson and Starr (1989). If we let E stand for error in predicting Y,

$$\nabla \equiv 1 - \frac{P(E|X = 1)}{P(E|X = 0)}$$

This amounts to a *measure* of the *degree* of necessity of X for Y, not a *test* of the *proposition* that X is necessary for Y. As such, it has two main disadvantages. First of all, even if one is able to overcome the considerable semantic obstacles to discussing degrees of necessity, ∇ cannot draw the line between a "near-necessary" condition and a necessary one. Here, we are only interested in the latter. Second, ∇ is insensitive to sample size: a zero in the ($Y = 1$, $X = 0$) cell produces the same results when the rest of the cells are filled with 1s that it would produce if they were filled with 1,000s. This is akin to producing regression coefficients without standard errors: the point estimates are useless without some way of measuring degrees of confidence.

A second measure used by the authors is the "scope" of a proposition, the "proportion of all cases not meeting the condition" (Anderson and McKeown 1987, 8). They include this measure as a way of weeding out "weak" necessary conditions—conceptually equivalent to our trivial necessary conditions—and in this capacity it serves well. Nevertheless, neither their measure of weakness nor our measure of trivialness serves as a test of necessity, and neither is intended to.

In sum, ∇ and scope are inappropriate for use in testing necessary condition hypotheses.

Testing for Necessity, Step One: Is *X* Necessary for *Y*?

The empirical evaluation of a necessary condition hypothesis requires two steps. The first involves determining

whether X is in fact a necessary condition for Y. If it is, the second involves determining whether X is trivially necessary. It is easy to confound these two questions, but it is crucial methodologically and conceptually to keep them distinct. A complete necessary condition analysis must include both, and the appropriate methodologies differ.

Deckard (1972, 217) takes a commonsensical approach to testing a necessary condition hypothesis: reject if a single counterexample can be found.

> [S]ince the hypothesis states that ideological homogeneity is not a necessary condition for delegate cohesion, one counterexample is sufficient to falsify the alternative hypothesis. (1972, 217)

As a general test, this approach suffers from four shortcomings. First, data are inevitably measured with error. This is true even in fields in which measurement is far more precise than even the most ambitious political scientist could hope to attain. Consider medical studies: human immunodeficiency virus (HIV) has been shown to cause acquired immune deficiency syndrome (AIDS), but the test for HIV occasionally produces false negatives, especially in the first 6–8 weeks following infection. People in whom HIV is mistakenly thought to be absent, of course, can and do develop AIDS. Therefore, a study that does not take the reliability of the HIV test into account would mistakenly conclude that HIV is not necessary for AIDS, simply because counterexamples can be found. This point should be relatively uncontroversial: many statistical procedures already exist that are designed to take it into account (see, e.g., Maddala 1992, chapter 11).

Second, reliability (the internal consistency of a measure) and validity (the extent to which the right phenomenon is being measured) are also a concern. In general, Carmines and Zeller suggest that "reliabilities should not be below .80 for widely used scales" (1979, 51). Social science practice typically conforms to this guideline; for example, Caldeira and Gibson report a Cronbach's alpha of 0.84, corresponding to an error rate of $1-\sqrt{0.84}$, or 8.35 percent, and constituting "a very strong indication of reliability" (1992, footnote 19). An error rate of that magnitude, however unproblematic in most applications, renders a necessary condition test that assumes perfect reliability useless. Similarly, if the data used suffer from problems of validity, the wrong phenomenon is being measured, so counterexamples are to be expected.

Even if all observations were made without any errors of measurement, reliability, or validity, the act of coding variables introduces what we call "grayness"—as, for example, when error is introduced by the dichotomization of a fundamentally nondichotomous phenomenon. For example, countries coded in the middle of the Polity III democracy scale include Chile in 1891, Guatemala in 1989, both Germany and Russia in 1917, and Japan in 1904. Few scholars would base strong conclusions about the behavior of democracies on any of these states, simply because they are marginal—but classification produces the illusion of precision.

Finally, the absence of counterexamples may be insufficient evidence of necessity if only a few cases have been examined. Clearly looking at one case, finding no counterexamples, and failing to reject the hypothesis of necessity is insufficient, but the "one counterexample falsifies" school of thought provides no guidance as to what constitutes a sufficient sample.

The practice of simply looking for counterexamples, however intuitively appealing, is therefore an unreasonable one to utilize, especially in large-N studies. The harder, and more interesting, question is how to test necessary condition hypotheses in real-world circumstances.

The p_l-Test

In principle, a necessary condition hypothesis states that the proportion of cases in cell $X = 0$ in the $Y = 1$ row should be zero. In practice, for reasons we have just discussed, the sample proportion of cases in that cell (call it $\hat{p}$) will depend on the error rate of the data. Hence, an appropriate test would be to reject the necessary condition hypothesis if the estimated error rate of the data is significantly lower than we would expect given $\hat{p}$. We therefore propose to test bivariate dichotomous necessary condition hypotheses by constructing a one-sided 95 percent confidence interval[3] around $\hat{p}$ and rejecting the necessary condition hypothesis if that confidence interval is inconsistent with estimates of error. More formally, the decision rule is:

> Reject the hypothesis of necessity if the lower bound of the one-sided 95 percent confidence interval around $\hat{p}$ is greater than the estimated error rate of the data.

[3] Solving $\sum_{j=T}^{n}\binom{n}{j}(p_L)^j(1-p_L)^{n-1}=\frac{\alpha}{2}$ and $\sum_{j=0}^{T}\binom{n}{j}(p_U)^j(1-p_U)^{n-1}=\frac{\alpha}{2}$, where $T=\sum_{j=1}^{N}x_j$, for p_L and p_U gives approximate $100(1-\alpha)\%$ lower and upper bounds; see Johnson, Kotz, and Kemp (1993, 129–132) for details. We choose 95 percent because it permits what we consider to be a reasonable number of counterexamples while substantially raising the bar (as the results will demonstrate). Those who desire a more or less stringent decision rule need only adjust α; no one should be a slave to the 5 percent rule.

This test guards against situations in which the number of counterexamples is high and the estimated error rate is low. For instance, if we were to witness five counterexamples out of twenty-five cases, the lower bound of the one-sided 95 percent confidence interval around $\hat{p}$ would be 0.0823, meaning that we know with 95 percent certainty that the population proportion is greater than 8.23 percent. Therefore, we would reject the hypothesis of necessity if the error rate of the data were (say) 3 percent or 5 percent, simply because there would be more counterexamples than could be accounted for by error. As the estimated error rate decreases and/or the number of counterobservations increases, the assertion that the counterobservations are the product of measurement error becomes less credible, and the test becomes harder to pass. We refer to this as the p_I-test because it guards against Type I error (rejecting a proposition when it is true.) This is the first stage of our test for dichotomous, bivariate necessary condition hypotheses.[4]

The problem, of course, is determining the error rate of the data given that we can only observe $\hat{X}$, not X. The simplest and most reasonable measurement model that includes an error term is $\hat{X} = X + \varepsilon$, where $\varepsilon \sim \left(0, \sigma_\varepsilon^2\right)$. This model assumes that on average the measure is accurate (i.e., ε has mean 0) and that ε is uncorrelated with X. Once ε has been included, $\hat{X}$ diverges from X, so simple observation of counterexamples cannot serve to falsify a necessary condition unless ε can be shown to be zero—a noble goal, but an almost inherently unattainable one. In the following section we examine this basic model in the context of the analysis of the reliability of indictors; subsequently we examine one method for estimating the variance of ε directly and use that estimate to draw conclusions about the necessary condition hypothesis.

Reliability

Estimates of reliability are uncommon: none of the studies we examine deals with the question in more than a cursory way. We therefore begin by demonstrating how it is possible to obtain an estimate of reliability and use it to determine whether or not the proportion of counterexamples observed is consistent with the hypothesis of low reliability. The most widely used estimate of reliability is Cronbach's alpha.[5] The procedure we outline below is simple: (1) obtain an estimate of reliability for X, (2) use that estimate to calculate the error rate of the data, and (3) apply the p_I-test to determine whether or not the implied error rate falls below the lower bound of a one-sided 95 percent confidence interval around $\hat{p}$. If so, the proportion of counterexamples is significantly higher than one would expect based on low reliability, and we reject the hypothesis of necessity.

We illustrate this procedure on data pertaining to the democratic peace from Russett (1995). These data suit our purpose because their reliability can be gauged. The measure of regime type, derived from the Polity project, is a combination of the Polity measures of democracy (*DEMOC*), autocracy (*AUTOC*), and power concentration (*PCON*): specifically, $REG = PCON \times (DEMOC - AUTOC)$. The measures *DEMOC* and *AUTOC* are themselves additive measures incorporating five variables that represent competitiveness of political participation, regulation of participation, competitiveness of executive recruitment, openness of executive recruitment, and executive constraints. The combination and weighting of these variables produces fairly reliable measures of both democracy (Cronbach's alpha equals 0.8207, implying an error rate of $1-\sqrt{alpha}$, or 9.41 percent) and autocracy (alpha equals 0.8408, implying an error rate of 8.31 percent).

Next, we estimate the standard error of measurement, or *SEM* (σ_ε^2, above). Following Dunbar (1998, 144), the formula is $SEM = SD_X \times \sqrt{1-alpha}$. The standard deviation of the democracy variable in the Polity II dataset used by Russett is 3.669, so the formula produces an *SEM* of 1.125. This means that if the observed democracy score equals 5 we can be 67 percent certain that the true democracy score lies between 3.875 and 6.125 and 95 percent certain that the true score lies between 2.75 and 7.25. The standard deviation of the autocracy variable is 3.518, producing an *SEM* of 1.014.

How much error does this imply for the composite *REG* variable? To answer this question we utilized Monte Carlo simulations. Two normally distributed variables, e_d and e_a, with mean zero and standard deviations of 1.125 and 1.014, respectively, were created as stand-ins for the uncertainty (ε) generated by the measures' imperfect reliability. These two variables were subtracted from the observed democracy and autocracy scores to produce "true" scores—that is, what the true democracy and autocracy scores would be if the simulated error corresponded to the actual error. A "true" regime score was then computed using the author's original formula. The

[4] Ragin (2000), who seeks to uncover not necessary conditions but rather "almost-necessary" ones, also bases his test on a binomial distribution; as is the case with Dion and the p_{II}-test (below), we take the convergence of methodologies among scholars approaching similar questions from different angles to be an encouraging sign.

[5] The seminal source is Cronbach (1951). For examples see Brady, Verba, and Schlozman (1995), Caldeira and Gibson (1992), Peffley and Hurwitz (1992), and Price and Zaller (1993). Though other examples can be found, Jones and Norrander are undoubtedly correct when they note that "political science research needs more of this kind of reliability assessment" (1996, footnote 3).

TABLE 2 A Test of the Democratic Peace Proposition

H_0 : Presence of. . .	. . . necessary for. . .	$k / n = \hat{p}$	p_I	Accept?
non-democracy	war	0/37=.000	.0000*	yes
non-democracy	use of force	8/237=.034	.0169*	yes
non-democracy	dispute	12/269=.045	.0259*	yes

Note: See Table 1 for hypotheses. Estimated error rate of 0.0513.

observed and "true" scores were dichotomized in accordance with the author's original coding rules[6] and cross-tabulated, and the procedure was repeated 1,000 times. On average, 2.57 percent of all cases were found to be nondemocracies miscoded as democracies, and 0.62 percent were found to be democracies miscoded as nondemocracies, implying an overall mean error rate of 3.19 percent.

Because the variable utilized in the study was *dyadic* democracy, however, the simple error rate would not suffice: it was necessary to estimate the probability that a conflictual nondemocratic dyad would be mistaken for a conflictual democratic dyad. There are three ways in which such a miscoding could occur: state i could be democratic and state j could be miscoded as democratic; i could be miscoded as democratic while j is genuinely democratic; and both i and j could be miscoded as democracies. The probability of either of the first two scenarios occurring, as mentioned above, is 2.57 percent, and the probability of the third's occurrence is $(2.57\%)^2$, or 0.07 percent. The probability that one of these three miscodings will occur is equal to one minus the product of the probabilities that they will not occur: $1 - (0.974 \times 0.974 \times 0.999)$, or 5.13 percent. This constitutes our estimate of error in the dichotomized, dyadic regime measure.

Armed with this information, we can now calculate the one-sided 95 percent confidence interval for $\hat{p}$ and determine whether or not our estimate lies within it. If not, we can conclude that error due to the measure's lack of reliability is not of sufficient magnitude to have produced the observed counterexamples. As stated above, the rule for the p_I-test is to reject the necessary condition hypothesis if the lower bound of the one-sided 95 percent confidence interval around $\hat{p}$ is greater than the estimated error rate of the data. As Table 2 demonstrates, the lower bound of the confidence intervals—0.00 percent for wars, 1.69 percent for use of force, and 2.59 percent for disputes—do not exceed 5.13 percent. We can therefore conclude that the data support the hypothesis of necessity, given what we are able to infer about the reliability of the data.

[6]See Russett (1993) for details.

In this case we need not address possible problems of validity and greyness or attempt to gauge the reliability of other variables used to calculate the composite score, simply because reliability is low enough to account for all of the observed counterexamples. If this were not so, it would be possible to adjust the allowable error rate in the p_I-test accordingly based on correlation with measures that reflect alternative conceptualizations or on the proportion of cases that fall into grey zones. Quite a few techniques exist for such purposes (see, e.g., Carmines and Zeller [1979] and Kmenta [1986]), and more can easily be derived as the need arises. An exegesis on the estimation of error rates is beyond the scope of this paper, however: the sources of error vary so widely that no set of prescriptions in an article of this length could be comprehensive. We seek only to establish and demonstrate the relevant test.

Measurement Error

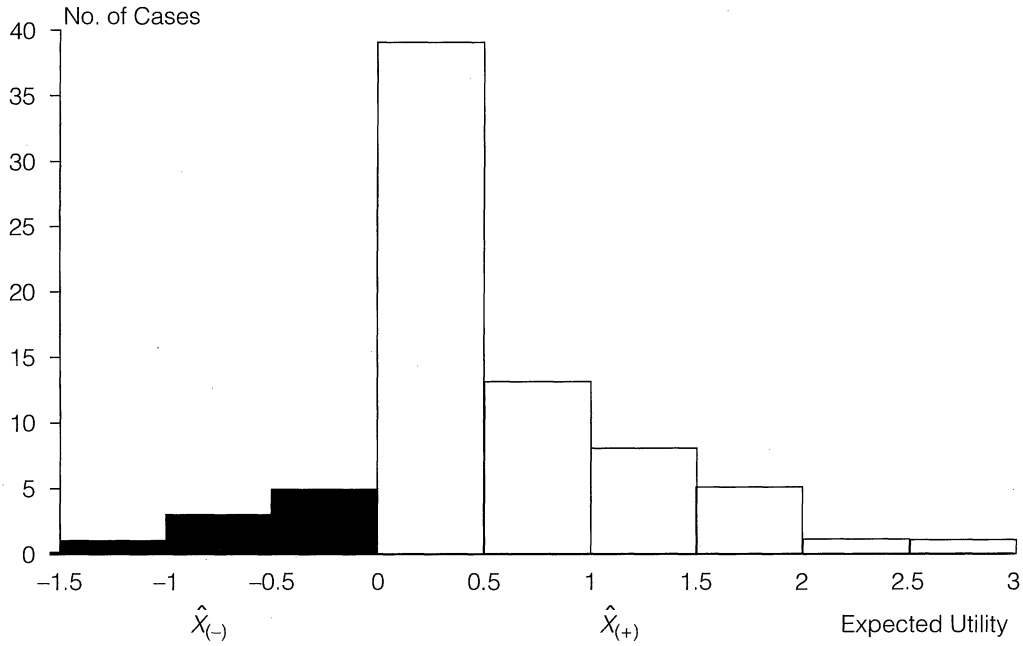
When measurement error is believed to be present it may be possible to model the error rate of the data and examine more closely the claim that it can explain counterexamples. If so, a suitable procedure can easily be devised.

Bueno de Mesquita's *War Trap* model provides excellent material for a test of necessity given measurement error: the author hypothesizes that "positive expected utility should be—within the bounds of measurement error—necessary but not sufficient for states to initiate conflict" (1981, 130). Bueno de Mesquita also hypothesizes that positive expected utility will be necessary for intervention and threats, but as the data for wars constitute the best evidence of the three for necessity,[7] only wars will be examined here.

Bueno de Mesquita essentially argues that positive expected utility (X) is a necessary condition for war initiation (Y) even though the measure of expected utility ($\hat{X}$) is negative for some wars. Again, rejecting the hypothesis based on a single counterexample would be unwise, but we cannot utilize the technique outlined above

[7]Bueno de Mesquita reports eleven countercases to his main hypothesis (129), but we were only able to locate nine in his Appendix C; we utilized the data from the appendix.

FIGURE 3 Measurement Error: Expected Utility of War Initiators



because expected utility is not an additive index. In this case, however, we have a continuous operationalization and a range hypothesis for the p_I-test. Accordingly, we need to reformulate the necessary-condition hypothesis slightly. In this case, necessity implies that $P(X < 0 | Y = 1) = 0$, which provides an opportunity not afforded us earlier: we can test the proposition that $p = 0$ net of measurement error.

Since we are dealing with a necessary condition we need only examine cases of war initiation ($Y = 1$). We divide these into two groups: those whose measured expected utility is negative, which we denote as $\hat{X}_{(-)}$, and those whose measured expected utility is positive, which we denote as $\hat{X}_{(+)}$. The data are illustrated in Figure 3.

What we need is an estimate of the variance of the measurement error, σ^2_ε. If Bueno de Mesquita is correct, all of the observations in $\hat{X}_{(-)}$ are actually nonnegative in X. These cases, represented by the shaded area in Figure 3, therefore contain critical information: we can use them to infer the measurement error variance.

The best-case scenario for the expected-utility hypothesis would be that all apparent countercases are equal to zero in reality. The most reasonable assumption in the absence of contradictory information is that all observations are measured with (on average) the same amount of error and that the error distribution is symmetric. If that is the case, and X is nonnegative, the variance of $\hat{X}_{(-)}$ around zero would have to equal the variance of the negative component of the error term, $\varepsilon_{(-)}$. Therefore, a conservative[8] estimate of σ^2_ε can be obtained by constructing a symmetrical distribution of eighteen observations centered at zero (with nine positive observations paralleling the nine negative observations that are believed to be the result of measurement error) and calculating the variance of that distribution.

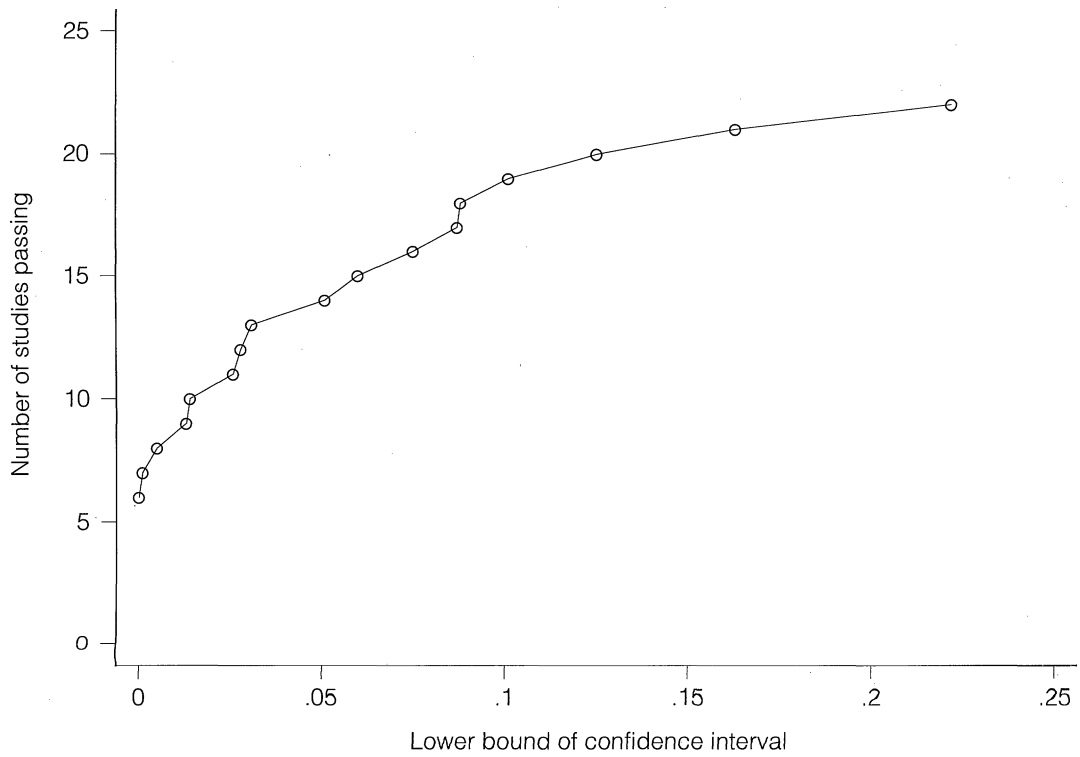
The variance of the data we observe (the $\hat{X}$s) is composed of two elements: the variance of the true values (σ^2_X) and the variance of the measurement error (σ^2_ε). Assuming that X and ε are independent, $\sigma^2_{\hat{X}} = \sigma^2_X + \sigma^2_\varepsilon$. Calculation based on the data presented in Appendix C yields:

$$\sigma^2_{\hat{X}} = \sigma^2_X + \sigma^2_\varepsilon$$
$$0.501 = -0.001 + 0.502$$

Obviously a value of –0.001 for the variance of σ^2_X makes no sense. If the variance of the measurement error were in fact this large (0.502), we would have to conclude that *none* of the variance in the $\hat{X}$s is due to the variation of the real Xs—in other words, real expected utility never

[8]In fact, very conservative: we have run a Monte Carlo simulation to determine the fairness of this estimate, and the ratio of our estimate of the variance to the actual variance of the (known) error term is on average almost exactly 1:3 over 1,000 trials. As our estimated variance approached but never exceeded the actual variance of the error term in any of those trials, we conclude that it constitutes the lowest reasonable estimate of σ^2_ε.

FIGURE 4 Sensitivity Test of Remaining Necessary Condition Hypotheses



varies. It is difficult to believe that the expected utility for all war initiators is the same. Thus by *reductio ad absurdum* we conclude that, even under assumptions most favorable to the theory, measurement error cannot account for all cases of negative expected utility and war initiation.

When Error Cannot Be Inferred

The question of error—whether due to problems of reliability, validity, or greyness—is central to the testing of necessary condition hypotheses but is usually ignored. As a result, none of the remaining studies from Table 1 contains information that would permit us to estimate it. In the absence of such information, the results of the p_I-test are difficult to interpret in any given case.

In the aggregate, however, the data provided by the other studies can help answer the larger question of whether any meaningful necessary conditions exist in social science research. We have calculated the one-sided 95 percent confidence interval for $\hat{p}$ in each of the studies listed, with the exception of those already tested, and in Figure 4 we illustrate the relationship between that lower bound—i.e., the maximum permissible error rate if the study were to pass—and the number of studies that would pass at that level.

The results should be sobering news to skeptics. Of the twnty-two remaining studies, six contained no counterexamples and would pass the p_I-test even if perfect measurement were assumed. If we assume an enviably low error rate (say, under 2 percent), another four studies would pass. If we assume an error rate of 5 percent, which we know to be reasonable based both on prior expectations of measurement error (e.g., standard criteria for acceptability using Cronbach's alpha) and on the example presented herein (the reliability of Russett's dichotomized democracy score), over half pass the test. Table 3 evaluates the necessary condition hypotheses for which estimates of error or reliability cannot be obtained. Overall, these findings demonstrate that necessary condition hypotheses cannot be rejected *a priori*; many pass the p_I- and p_{II}-tests under plausible assumptions about measurement error and reliability.

Is an Absence of Counterexamples Enough? The p_{II}-Test

As mentioned above, a paucity of counterexamples can be insufficient evidence of necessity if the number of observations is quite small. In statistical terms, we need to guard against Type II error (failing to reject the proposi-

TABLE 3 An Empirical Test of Some Necessary Condition Hypotheses

Author	H_0 : Presence of . . .	. . . necessary for. . .	$k / n = \hat{p}$	p_I	p_{II}
Anderson and McKeown	activation	war	5/77=.06	.026[b]	.00*
Anderson and McKeown	attention I	war	12/71=.17	.101	.00*
	attention II	war	13/64=.20	.125	.00*
	attention III	war	19/59=.32	.222	.00*
	attention IV	war	2/72=.03	.005[a]	.00*
	attention V	war	5/40=.12	.051	.00*
	attention VI	war	11/72=.15	.088	.00*
	attention VII	war	11/40=.27	.163	.00*
Diamond	wealth	democracy	3/62=.05	.013[a]	.00*
	development	democracy	1/65=.02	.001[a]	.00*
Goertz and Diehl	shock	rivalry termination	2/26=.08	.014[a]	.00*
	shock	rivalry initiation	6/45=.13	.060	.00*
Kugler and Organski	power transition	war	0/5=.00	.000*	.10
Osherenko and Young	leadership	institutions	0/5=.00	.000*	.10
Ostrom	design principle	institutions	0/8=.00	.000*	.03*
Rosenthal	Modérés	R.P.F. alliance	2/13=.15	.028[b]	.01*
Ryan	cc/GATT utility	USTR action	0/37=.00	.000*	.00*
Schweller	decl. dem.	no prev. war	0/10=.00	.000*	.02*
Schweller	decl. non-dem.	prev. war	0/20=.00	.000*	.00*
Siverson and Starr	opp./will.	war	44/455=.10	.075	.00*
major powers	opp./will.	war	9/152=.06	.031[b]	.00*
minor powers	opp./will.	war	35/303=.12	.087	.00*

Legend: *Do not reject. [a]Pass assuming ≤ 2% error. [b]Pass assuming ≤ 5% error.

See Table 1 for hypotheses. p_I refers to lower bound of 95% confidence interval, one-sided test; numbers for p_{II} -tests are probability of Type II error given N, $H_1 : p = 0.5$, $\alpha = 0.05$.

tion when it is false.) Few social-science endeavors take the possibility of Type II error into account, but we must. Imagine, for example, an author who presented only a single confirming case as evidence for a theory. $\hat{p}$ would equal 0.00, and even if we assumed perfect measurement the data would pass the p_I-test, despite the fact that there are too few observations to warrant the conclusion of necessity. The tests outlined above are not sensitive to this problem. We therefore need to establish a criterion for deciding whether the probability of Type II error is large enough to warrant rejecting the hypothesis of necessity.

A standard power test is designed to do exactly that. Power, which is equal to one minus the probability of Type II error, is a function of α (the significance level used in the p_I-test; we follow convention by setting α = 0:05), the number of cases N, and the values of p posited by the null and alternative hypotheses (see *inter alia* Neter, Wasserman and Whitmore 1993, 339–348.) In the spirit of classical multivariate hypothesis testing, which posits no relationship between two variables (e.g., H_0: β = 0 in regression, probit, etc.) and requires that such a hypothesis be rejected with a high degree of certainty, it seems reasonable to posit a uniform distribution of X (which, here, would imply $p = 0.50$) and to require that such a supposition be rejected with a high degree of certainty. (The logic of Bayesian hypothesis testing, which usually posits a uniform distribution, also suggests this value as the baseline reference point.) Given the common choice of an α-level of 0.05 we do the same for the β-level.

We therefore propose as a test the requirement that the data used in the p_I-test contain sufficient observations to distinguish between data consistent with necessity and data consistent with a uniform distribution. This criterion rejects the hypothesis of necessity for small sample sizes even if $\hat{p}$ is very low. We refer to this as the p_{II}-test. Based on this criterion, we would reject all hypotheses with a sample size below seven due to the risk of Type II error, even absent counterexamples.[9] Here the rule becomes:

Reject the hypothesis of necessity if $n < 7$.

[9] Testing $p = 0.50$ against $p = 0.05$ comports with logic and experience, but adjusting for different estimates of either is simple: the formula is $n = \left[\frac{z_{1-\alpha}\sqrt{p_0(1-p_0)} + z_{1-\beta}\sqrt{p_A(1-p_A)}}{p_A - p_0}\right]^2$, where $z_{1-\alpha}$ refers to the $(1 - \alpha)$ quantile of the normal distribution (analogously for β) and p_0 and p_A refer to proportions hypothesized under null and alternative hypotheses (see Pagano and Gauvreau 1993). Many statistics programs make these calculations trivial (see e.g. Stata's `sampsi` command).

Interestingly, this corresponds closely to Dion's (1998) requirement, based on Bayesian methods and 95 percent confidence intervals, of a sample size of six (see also Ragin 1999.) Table 3 gives our analyses of the remainder of our necessary condition hypotheses using the p_I- and p_{II}-tests. The final column lists the probability of Type II error under these assumptions.

Summary

Downs and others have expressed skepticism regarding the existence of necessary conditions. In our analysis we find that, out of twenty-six total hypotheses tested, seven pass unambiguously, and another seven show considerable promise: although error rates cannot be inferred, their distributions are consistent with the hypothesis of necessity and measurement error of up to 5 percent. At the same time, our tests clearly raise the bar: all of the necessary condition hypotheses listed were accepted in the original studies. We conclude both that a large number of theoretically important necessary condition hypotheses exist and that a substantial proportion can be supported empirically.

Testing for Necessity, Step Two: Evaluating Trivialness

Step one of necessary condition methodology determines if X is necessary for Y. If this is not the case, step two becomes irrelevant. Hereinafter the discussion assumes that X is a necessary condition for Y: counterexamples may exist, but not in sufficient number to invalidate the necessary condition hypothesis, and the number of observations is high enough to inspire confidence.

First of all, we should emphasize that trivialness is an empirical (rather than theoretical or logical) concept. We do not judge the theoretical merits of the necessary condition hypotheses we have included: really trite necessary-condition hypotheses just do not make it into print. Critics of necessary-condition hypotheses provide the main exceptions to this rule by providing trivial hypotheses as an argument against the existence of meaningful necessary conditions in general. This practice makes no more sense than arguing against the correlation coefficient by providing examples of nonsignificant correlations. Rather than take *a priori* positions, the reasonable thing to do is evaluate the trivialness of the hypothesis empirically. To do so, we need a statistical test of trivialness.

As far as we are aware, no one has formalized the concept of a trivial necessary condition. To do so, we examine the underlying principle of examples of trivial necessary conditions. To take the Downs quote (see above) as an example, what makes gravity trivially necessary for war? The answer is that it is impossible to find armies—at least at this point in history—that operate without gravity. In all cases, war or no-war, gravity is present: X is constant for all values of Y. In statistical terms, one of the variables does not vary.

Again, the democratic peace literature provides a useful illustration. Democratic peace theory posits that at least one nondemocracy in a dyad is a necessary condition for war. In the eighteenth century, a time when by most measures the number of democracies could be counted on the fingers of one hand, this was not a terribly important necessary condition for war—by far the majority of dyads fulfilled it. In the late twentieth century, however, the proliferation of democracy has made it very important. This can be an issue for research designs like Schweller's (1992) examination of power transitions, democracy, and war. In Schweller's no-dual-democracy/war category, twelve out of his seventeen cases are from before 1800. If we were to partition his results into pre-1800 and post-1800 power transitions, we would find in both tables that no democratic power transition wars occurred; in the pre-1800 table, we would also notice that there were no democratic power transitions in which war *could* have occurred. Empirically, the presence of at least one nondemocracy in a dyad was a trivial necessary condition for power-transition war before 1800 because democracy did not vary.

The democratic peace literature also illustrates the fact that trivialness can be due to an absence of variation in the *dependent* variable: nondemocracy can be trivially necessary for war in a period in which there are no (or very few) wars. In her critique of the democratic peace literature, Gowa (1999, 62–63) concludes that regime type has no effect on the probability of war between 1919 and 1938, a period in which quite a few democratic dyads existed. Though it may seem paradoxical, this does not mean that the hypothesis that nondemocracy is necessary for war has been falsified: the reported probability of war among democratic states was 0.0000. Rather, war was exceptionally rare in the period as a whole (the probability of war among nondemocratic dyads was 0.0003), making it virtually impossible to use that particular sample to assess the relationship between regime type and war. The decision to exclude 1914–1918 and 1939–1945 from the interwar period, as well as from any prior or subsequent periods, exacerbates this problem: wars are rare events to begin with, and splitting the sample into different periods and removing two wars that spanned the globe increases trivialness dramatically. Prior to 1800 in Schweller's study, nondemocracy was trivially necessary for war because there were virtually no democracies; Gowa's findings re-

flect the fact that nondemocracy was trivially necessary for war because the sample contained very few wars.

The basic principle for the examination of the trivialness claim consists in comparing the "region of necessity" with the nonnecessary condition range of X and Y values. For example, in our standard 2×2 case the region of necessity is the row $Y = 1$, so testing for trivialness revolves around comparing that row to the data in the $Y = 0$ row. Trivialness takes different forms depending on the character of the data and the nature of necessary-condition hypothesis.

Investigating trivialness requires a comparison of the region of necessity with the cases that were excluded in step 1. The upper left-hand table in Figure 5 illustrates what happens in the case in which there is no variation in X: the two rows have identical proportions. Both the gravity and the pre-1800 democratic peace examples show this kind of data configuration. This represents an archetypal example of trivial necessity. The upper right-hand table illustrates the absence of variation in Y of the sort that drives Gowa's null results.

The bottom table in Figure 5, by contrast, shows an example of a nontrivial necessary condition. Unlike either example, we have variation in X as well as Y. We see that the distribution of data in the nonnecessity region is quite different from that in the region of necessity. As the number of cases in either the $X = 0/Y = 0$ cell or the $X = 1/Y = 1$ cell decreases toward zero, the hypothesis becomes more trivial.

We need to determine whether or not the variation in one row or column is significantly different from the variation in the other. To do so, we simply compare the distributions across rows and columns. If the distributions differ significantly, X is nontrivial. This procedure establishes trivialness as a statistical, rather than a substantive, concept.

At this point, we will refer to the upper-left table in Figure 5, since we have chosen to test data in the $Y = 1$ row. If the proposed necessary condition is trivial, the variation in the $Y = 0$ row will be a plausible result of the process that produced variation in the $Y = 1$ row. If the proposed necessary condition is not trivial, however, there should be variation in the $Y = 0$ row beyond what could reasonably be expected given what we already know about the $Y = 1$ row.[10]

Nontrivialness therefore requires a *significant* difference between the distributions in the $Y = 0$ and $Y = 1$ rows. The χ^2 test of homogeneity fits our needs exactly, as it tests for differing distributions of cases over the rows of a 2×2 table. If the distribution in the $Y = 0$ row significantly differs from the distribution in the $Y = 1$ row, the proportion of $X = 0$ cases in the $Y = 0$ row must be significantly different from the $Y = 1$ one, and we can conclude that the condition is not trivial.

FIGURE 5 Trivialness and Necessity

Y \ X	0	1
0	0	100
1	0	100

Trivially necessary

Y \ X	0	1
0	100	100
1	0	0

Trivially necessary

Y \ X	0	1
0	100	100
1	0	100

Non-trivially necessary

We now consider the trivialness of some of the necessary condition hypotheses analyzed above. It should be emphasized that not all 2×2 tables with a significant χ^2 correspond to nontrivial necessary conditions: the χ^2 test is *only* a valid test of trivialness if the hypothesis has passed the first hurdle. Therefore, we eliminate any studies that failed the p_{II}-test; in addition, we only retain studies that either passed the p_I-test unambiguously or proved to be plausible candidates (specifically, those in which reliability could not be ascertained but findings were consistent with necessity and a low ($\leq$ 5 percent) error rate). Table 4 presents the results of this analysis.[11] We see that none of these studies is trivial in the empirical sense: the χ^2s are all significant at the .05 level or above.[12]

[10]In the case of continuous data, the analogous procedure would be to compare the Y-intercept to the regression line, which indicates the character of the X–Y relationship outside the $X = 0$, $Y = 0$ region. If the regression estimate for X is not statistically significant then there is no evidence that the necessity and nonnecessity regions differ: hence, we have a trivial necessary condition.

[11]We inferred that the number of cases in the "no war/absent treatment" cell of Siverson and Starr's analysis of "Any-WBN/Any-WAP" cases was 525, based on the cell proportions. See Table 7 of Siverson and Starr (1989, 40).

[12]One might be concerned about the χ^2 test in some cases since the test can give misleading results when some cells have zero or small values. This of course is exactly what happens with trivial necessary conditions. Accordingly, we confirmed the χ^2 analysis using Fisher's exact test, which is not subject to these problems. The results were very close to the χ^2 ones. When any cell contained fewer than three observations, we reported p based on Fisher's exact test rather than on χ^2. (See, e.g,. Hays [1973, 735–740] for details regarding these tests.)

TABLE 4 An Empirical Test of the Trivialness of Some Necessary Condition Hypotheses

Author	H_0 : Presence of . . .	. . . necessary for . . .	χ^2	p	Trivial?
Anderson and McKeown	activation	war	8.9	0.00	no
	attention IV	war	455.1	0.00	no
Diamond	wealth	democracy	36.7	0.00	no
	development	democracy	24.9	0.00	no
Goertz and Diehl	shock	rivalry termination	4.2	0.04	no
Ostrom	design principle	institutions	—	—	†
Rosenthal	Modérés	R.P.F. alliance	7.9	0.01	no
Russett	non-democracy	war	6.0	0.02	no
	non-democracy	use of force	25.7	0.00	no
	non-democracy	dispute	24.0	0.00	no
Ryan	cc/GATT utility	USTR action	40.0	0.00	no
Schweller	decl. dem.	no prev. war	30.0	0.00	no
	decl. non-dem.	prev. war	30.0	0.00	no
Siverson and Starr*	opp./will.	war	230.2	0.00	no

† insufficient $Y = 0$ cases to calculate

* major powers.

Once again, critiques based on trivialness seem unfounded. In addition to trivial necessary conditions like gravity and water, there appear to be nontrivial necessary condition in political affairs. While all these hypotheses merit further examination, we cannot with a wave of the hand dismiss them as trivial.

Necessity When X and Y Are Continuous

The vast majority of necessary condition analyses, as the above passages indicate, are carried out on binary data. Occasionally, however, researchers wish to test a hypothesis of necessity on continuous data, for which our p_I-test would be inappropriate.

We know of only one such necessary condition hypothesis in the international relations literature that has this character: the relation between uncertainty and war. A number of authors, such as Stoessinger (1990) and Blainey (1973), have argued that war serves as an information-gathering device to determine which side is more powerful. In a situation of perfect and complete information, war would no longer be necessary since the weaker side would know it is weaker (analogously for the stronger side) and a negotiated settlement reflecting relative power would ensue. If the X-axis represents uncertainty (uncertainty increases with X) and the Y-axis the number of wars, the necessary condition hypothesis states that when $X = 0$ (no uncertainty), $Y = 0$ (no wars).

Bueno de Mesquita and Lalman (1992, 60–65) are unique in that they operationalize both uncertainty (UNC) and war and test this proposition. Their use of logit analysis is quite natural given the dichotomous character of their dependent variable but is not appropriate for our test of necessity: logit has an S-shaped functional form designed to keep predicted values between zero and one. This is exactly what we do not want. The fact that regression lines can pass through the origin is a desirable characteristic for testing necessity. It therefore serves as a useful exercise to reanalyze their data (which they have kindly provided). Our reestimation of their model—given in Table 5—shows that we can reject the hypothesis that the Y-intercept is zero or less with greater than 99 percent certainty. We also estimated a model that included UNC^2, in case a curvilinear model proved to be a more accurate representation of the data, but the negative coefficient for the UNC term and the positive coefficient for the UNC^2 term indicate a convex, not concave, regression curve, and the Y-intercept increased from 0.628 to 0.870 on a scale of 0 (no war) to 1 (war). We conclude, as they did, that their evidence does not support the hypothesis that uncertainty is a necessary condition for war.

Conclusions

This essay has served to demonstrate that necessary conditions are neither rare in, nor inapplicable to, political research. We were prompted to write it by the discrepancy between the prevalence of implicit or explicit necessary condition hypotheses and the dearth of literature regarding their implications for empirical research. We have demonstrated the need to evaluate both necessity and trivialness. We have developed an array of techniques tailored to these tasks and have demonstrated how they can

TABLE 5 Asymptotic Testing of Necessity: War and Complete Information

Variable	Coefficient	Std. Err.	*t*-stat	$p > \|t\|$
UNC	−0.122	0.148	−0.826	0.409
CONSTANT	0.628	0.037	17.198	0.000
UNC	−3.004	0.725	−4.142	0.000
UNC^2	6.053	1.492	4.057	0.000
CONSTANT	0.870	0.070	12.479	0.000

Original analysis: Bueno de Mesquita and Lalman (1992, 65).

be utilized by reanalyzing existing studies. The results make a strong case for the existence of meaningful necessary condition relationships in political science.

We have also attempted to lay some of the groundwork for future research utilizing necessary conditions. Because necessary conditions fly in the face of much of conventional political science practice, our focus has primarily been on describing their most basic empirical implications. If the practice and recognition of necessary-condition hypothesizing become more widespread, a richer and more detailed dialogue will no doubt develop.

References

Agresti, Alan. 1990. *Categorical Data Analysis.* New York: John Wiley & Sons.

Anderson, Paul, and Timothy McKeown. 1987. "Changing Aspirations, Limited Attention, and War." *World Politics* 40:1–29.

Blainey, Geoffrey. 1973. *The Causes of War.* New York: Free Press.

Brady, Henry, Sidney Verba, and Kay Lehman Schlozman. 1995. "Beyond SES: A Resource Model of Political Participation." *American Political Science Review* 89:271–294.

Bueno de Mesquita, Bruce. 1981. *The War Trap.* New Haven: Yale University Press.

Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason: Domestic and International Imperatives.* New Haven: Yale University Press.

Caldeira, Gregory A., and James L. Gibson. 1992. "The Etiology of Public Support for the Supreme Court." *American Journal of Political Science* 36:635–664.

Carmines, Edward G., and Richard A. Zeller. 1979. "Reliability and Validity Assessment." Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–017. Beverly Hills and London: Sage Publications.

Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16:297–333.

Deckard, Barbara. 1972. "State Party Delegations in the U.S. House of Representatives–A Comparative Study of Group Cohesion." *Journal of Politics* 34:199–222.

Diamond, Larry. 1992. "Economic Development and Democracy Reconsidered." In *Reexamining Democracy: Essays in Honor of Seymour Martin Lipset,* ed. Gary Marks and Larry Diamond. Newbury Park, CA: Sage Publications.

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

Downs, George. 1989. "The Rational Deterrence Debate." *World Politics* 41:225–237.

Doyle, Michael. 1983. "Kant, Liberal Legacies, and Foreign Affairs." *Philosophy & Public Affairs* 12:205–235, *Philosophy & Public Affairs* 12:323–353.

Dunbar, George L. 1998. *Data Analysis for Psychology.* New York: St. Martin's Press.

George, Alexander L., and Richard Smoke. 1974. *Deterrence in American Foreign Policy: Theory and Practice.* New York: Columbia University Press.

George, Alexander L., and Richard Smoke. 1989. "Deterrence and Foreign Policy." *World Politics* 41:170–182.

Goertz, Gary, and Paul Diehl. 1995. "The Initiation and Termination of Enduring Rivalries: The Impact of Political Shocks." *American Journal of Political Science* 39:30–52.

Gowa, Joanne. 1999. *Ballots and Bullets: The Elusive Democratic Peace.* Princeton: Princeton University Press.

Hays, William L. 1973. *Statistics for the Social Sciences,* 2nd ed. New York: Holt, Rinehart and Winston, Inc.

Hildebrand, David K., James D. Laing, and Howard Rosenthal. 1976. "Prediction Analysis in Political Research." *American Political Science Review* 70:509–535.

Johnson, Norman L., Samuel Kotz, and Adrienne W. Kemp 1993. *Univariate Discrete Distributions.* New York: John Wiley and Sons.

Jones, Bradford S., and Barbara Norrander. 1996. "The Reliability of Aggregated Public Opinion Measures." *American Journal of Political Science* 40:295–309.

Kmenta, Jan. 1986. Elements of Econometrics, 2nd ed. London: Macmillan.

Kugler, Jacek, and A. F. K. Organski. 1989. "The Power Transition: A Retrospective and Prospective Evaluation." In *Handbook of War Studies,* ed. Manus Midlarsky. Ann Arbor: University of Michigan Press.

Lebow, Richard Ned. 1981. *Between Peace and War: The Nature of International Crisis.* Baltimore: Johns Hopkins University Press.

Maddala, G. S. 1992. *Introduction to Econometrics.* New York: Macmillan.

Majeski, Stephen, and David Sylvan. 1984. "Simple Choices and Complex Calculations: A Critique of The War Trap." *Journal of Conflict Resolution* 28:316–340.

Most, Benjamin, and Harvey Starr. 1989. *Inquiry, Logic, and International Politics.* Columbia: University of South Carolina Press.

Neter, John, William Wasserman, and G. A. Whitmore 1993. *Applied Statistics*, 4th ed. Englewood Cliffs, N.J.: Prentice Hall.

Osherenko, Gail, and Oran Young. 1993. "The Formation of International Regimes: Hypotheses and Cases." In *Polar Politics: Creating International Environmental Regimes*, ed. Oran Young and Gail Osherenko . Ithaca: Cornell University Press.

Ostrom, Elinor. 1991. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge: Cambridge University Press.

Pagano, Marcello, and Kimberlee Gauvreau. 1993. *Principles of Biostatistics.* Belmont, Calif: Duxbury Press.

Peffley, Mark, and Jon Hurwitz. 1992. "International Events and Foreign Policy Beliefs: Public Response to Changing Soviet-U.S. Relations." *American Journal of Political Science* 36:431–461.

Price, Vincent, and John Zaller. 1993. "Who Gets The News? Alternative Measures of News Reception and Their Implications for Research." *Public Opinion Quarterly* 57:133–164.

Ragin, Charles. 2000. *Fuzzy-Set Social Science.* Chicago: University of Chicago Press.

Ragin, Charles. 1999. "Conceptualizing Complexity: A Fuzzy-Set Approach." Presented at the Annual Meeting of the American Political Science Association, Atlanta, GA.

Rosenthal, Howard. 1969. "The Electoral Politics of Gaullists in the Fourth French Republic: Ideology or Constituent Interest?" *American Political Science Review* 63:476–487.

Russett, Bruce. 1995. "The Democratic Peace: 'And Yet It Moves.'" *International Security* 19:164–175.

Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post-Cold War World.* Princeton: Princeton University Press, 1993.

Ryan, Michael P. 1995. "USTR's Implementation of 301 Policy in the Pacific." *International Studies Quarterly* 39:333–350.

Schweller, Randall. 1992. "Domestic Structure and Preventive War: Are Democracies More Pacific?" *World Politics* 44:235–269.

Simowitz, Roslyn, and Barry L. Price. 1990. "The Expected Utility Theory of Conflict: Measuring Theoretical Progress." *American Political Science Review* 84:439–460.

Siverson, Randolph M., and Harvey Starr. 1989. "Alliance and Border Effects on the War Behaviour of States: Refining the Interaction Opportunity Model." *Conflict Management and Peace Science* 10:21–46.

Stoessinger, John. 1990. *The Might of Nations: World Politics In Our Times.* New York: McGraw-Hill.